(FILE 'HOME' ENTERED AT 10:56:00 ON 09 APR 2002)

FILE 'USPATFULL' ENTERED AT 10:57:11 ON 09 APR 2002
L1 44467 S (PROTEIN OR POLYPEPTIDE OR PEPTIDE)/CLM,TI
L2 299 S 702/19/NCL
L3 73 S L2 AND L1

=> d an,pi,pn,ti,ab,clm 1,2,5,7,19,20,22,26,30,42,61,62,63,65,66,70,71

L3 ANSWER 1 OF 73 USPATFULL
AN 2002:73134 USPATFULL
PI US 2002040275 A1 20020404
PI US 2002040275 A1 20020404
TI Proteomic analysis
AB The present invention provides methods for analyzing proteomes, as cells or lysates. The analysis is based on the use of probes that have specificity to the active form of proteins, particularly enzymes and receptors. The probes can be identified in different ways. In accordance with the present invention, a method is provided for generating and screening compound libraries that are used for the identification of lead molecules, and for the parallel identification of their biological targets. By appending specific functionalities and/or groups to one or more binding moieties, the reactive functionalities gain binding affinity and specificity for particular proteins and classes of proteins. Such libraries of candidate compounds, referred to herein as activity-based probes, or ABPs, are used to screen for one or more desired biological activities or target proteins.
CLM What is claimed is:
1. A method for screening for the bioactivity of a candidate compound toward a group of related target **proteins** in a proteomic mixture of **proteins** from a cell, employing at least one probe, each probe characterized by comprising a reactive functionality group specific for said group of target **proteins** and a ligand, each probe of the formula: R(F--L)--X wherein: X is a ligand for binding to a reciprocal receptor and/or providing a detectable signal; L is an alkylene, oxyalkylene or polyoxyalkylene linking group, wherein said oxyalkylenes are of from 2 to 3 carbon atoms; F is a phosphonate or sulfonyl functional group reactive at an active site of a target enzyme; and R is bonded to F and a moiety of less than 1 kDal providing specific affinity for said enzymes, and when F is phosponate, F is fluorine and when F is sulfonyl, R is an aryl or heteroaryl group; said method comprising: combining at least one probe with an untreated portion of said mixture and with a portion inactivated with a non-covalent agent under conditions for reaction with said target **proteins**; sequestering **proteins** conjugated with said at least one probe from each of said mixtures; determining the **proteins** that are sequestered; and comparing the amount of each of the **proteins** sequestered from the untreated portion and the inactivated portion as indicative of the bioactivity of said candidate compound with said target **proteins**.

2. A method according to claim 1, wherein said probe is a fluorophosphonyl and said enzymes are serine hydrolases.

3. A method according to claim 1, wherein said probe is a sulfonate, R is a heteroaryl and said enzymes are aldehyde dehydrogenases.

4. A method according to claim 3, wherein said heteroaryl is pyridyl.

5. A method according to claim 1, wherein X is biotin.

6. A method according to claim 1, wherein said non-covalent agent is heat.

7. A method for screening for the bioactivity of a candidate compound toward a group of related target **proteins** in a proteomic mixture of **proteins** from a cell, employing at least one probe, each probe characterized by comprising a reactive functionality group specific for said group of target **proteins**, a ligand and having other than the natural isotope distribution of at least one element, each probe of the formula: R (F--L)--X wherein: X is a ligand for binding to a reciprocal receptor and/or providing a detectable signal; L is an alkylene, oxyalkylene or polyoxyalkylene linking group, wherein said oxyalkylenes are of from 2 to 3 carbon atoms; F is a phosphonate or sulfonyl functional group reactive at an active site of a target enzyme; and R is bonded to F and a moiety of less than 1 kdal providing specific affinity for said enzymes, and when F is phosphonate, F is fluorine and when F is sulfonyl, R is an aryl or heteroaryl group; said method comprising: combining at least one probe with an untreated portion of said mixture and with a portion inactivated with a non-covalent agent under conditions for reaction with said target **proteins**; sequestering **proteins** conjugated with said at least one probe from each of said mixtures; determining the **proteins** that are sequestered and the probe by mass spectrometry; and comparing the amount of each of the **proteins** sequestered from the untreated portion and the inactivated portion as indicative of the bioactivity of said candidate compound with said target **proteins**.

8. A method according to claim 7, wherein the unnatural isotope is hydrogen, carbon or nitrogen.

9. A method for determining in a proteomic mixture the presence of active target members of a group of related **proteins**, said related **proteins** related in having a common functionality for conjugation at an active site, employing a probe comsaid method comprising: combining said proteomic mixture in wild-type form with a probe comprising a fluorophosphonate or sulfonate reactive functionality specific for said active site when active, under conditions for conjugation of said probe to said target members; combining said proteomic mixture after non-specific deactivation with said probe under said same conditions; determining the presence of target members conjugated with said probe in said proteomic mixtures in active and inactive form; whereby when said target members are conjugated to target members in said proteomic mixture in active form and in less amount in inactive form, the presence of active members is determined.

10. A method according to claim 9, wherein said probe comprises a detectable lable.

11. A method according to claim 9, wherein said proteomic mixture is the composition from an intact cell.

12. A method for determining in a plurality of proteomic mixtures the presence of active target members of a group of related **proteins** in each of said proteomic mixtures, said related **proteins** related in having a common functionality for conjugation at an active site, said method comprising: combining each of said proteomic mixtures in wild-type form with a probe comprising a reactive fluorophosphonates or sulfonate functionality specific for said active site when active, under conditions for conjugation of said probe to said target members; determining the presence of target members conjugated with said probe in said proteomic mixtures; analyzing for the presence of target members

conjugated with said probe using simultaneous individual capillary electrokinetic analysis or capillary HPLC; whereby when said target members are conjugated to target members in said proteomic mixtures, the presence of active target members is determined.

13. A method for determining in a plurality of proteomic mixtures the presence of active target members of a group of related **proteins** in each of said proteomic mixtures, said related **proteins** related in having a common functionality for conjugation at an active site, said method comprising: combining each of said proteomic mixtures in wild-type form with a probe comprising a fluorophosphonate or sulfonate reactive functionality specific for said active site when active, under conditions for conjugation of said probe to said target members; determining the presence of target members conjugated with said probe in said proteomic mixtures; analyzing for the presence of target members conjugated with said probe using simultaneous individual capillary electrokinetic analysis or capillary HPLC; whereby when said target members are conjugated to target members in said proteomic mixtures, the presence of active target members is determined.

14. A method according to claim 13 including the additional steps of: inactivating a portion of said protemic mixture; combining said inactivated proteomic mixture with said probe under conditions for conjugation; analyzing for the presence of target members conjugated with said probe in said inactivated proteomic mixture; and rejecting conjugates from said wild-type proteomic mixture in less amount than the amount of conjugate from said inactivated mixture.

15. A method for determining in a plurality of proteomic mixtures the presence of active target members of a group of related **proteins** in each of said proteomic mixtures, said related **proteins** related in having a common functionality for conjugation at an active site, said method comprising: combining each of said proteomic mixtures in wild-type form with a probe comprising a sulfonate aryl or heteroaryl A method for determining in a plurality of proteomic mixtures the presence of active target members of a group of related **proteins** in each of said proteomic mixtures, said related **proteins** related in having a common functionality for conjugation at an active site, said method comprising: combining each of said proteomic mixtures in wild-type form with a probe comprising a fluorophosphonate or sulfonate reactive functionality specific for said active site when active, under conditions for conjugation of said probe to said target members; determining the presence of target members conjugated with said probe in said proteomic mixtures; analyzing for the presence of target members conjugated with said probe using simultaneous individual capillary electrokinetic analysis or capillary HPLC; whereby when said target members are conjugated to target members in said proteomic mixtures, the presence of active target members is determined. reactive functionality specific for said active site when active, under conditions for conjugation of said probe to said target members; determining the presence of target members conjugated with said probe in said proteomic mixtures; analyzing for the presence of target members conjugated with said probe using simultaneous individual capillary electrokinetic analysis or capillary HPLC; whereby when said target members are conjugated to target members in said proteomic mixtures, the presence of active target members is determined.

L3 ANSWER 2 OF 73 USPATFULL
AN 2002:67530 USPATFULL
PI US 2002038184 A1 20020328
PI US 2002038184 A1 20020328

TI Methods and systems for estimating binding affinity

AB This application discloses methods and systems of predicting binding affinity between a ligand and a receptor. In one embodiment, the predicted binding affinity (pK.sub.i) is determined by at least using a formula pK.sub.i=C0+C1*vdW+C2*Att_pol+C3*(Att_pol*Att_pol+Rep_pol*Rep_pol). vdW represents the van der Waals interaction energy between the ligand and the receptor. Att_pol represents the surface area of the ligand forming complimentary polar interactions with the receptor. Rep_pol represents the surface area of the ligand forming uncomplimentary polar interactions with the receptor. This application also discloses an improved process of calculating linear interpolation of grid-based vdW energy. A first non-linear function is transformed into a less non-linear second non-linear function to reduce the error in linear interpolation. A trilinear interpolation process is applied to the second non-linear function. The value obtained is reverse transformed to produce an estimated vdW energy.

CLM What is claimed is:

1. A method of estimating a binding affinity between first and second interacting molecular entities, said method comprising: defining at least one surface area descriptor of the interaction, said descriptor comprising an amount of non-neutral surface area of the first molecular entity that is proximate to a non-neutral portion of the second molecular entity, and using said amount of non-neutral surface area of the first molecular entity in a formula for numerically estimating said binding affinity.

2. The method of claim 1, comprising defining an amount of surface area of the first molecular entity having a first charge polarity that is proximate to a portion of the second molecular entity having the same charge polarity.

3. The method of claim 2, comprising defining an amount of surface area of the first molecular entity having a first charge polarity that is proximate to a portion of the second molecular entity having the opposite charge polarity.

4. The method of claim 1, additionally comprising calculating a van der Waals interaction energy between said first molecular entity and said second molecular entity.

5. The method of claim 1, comprising using said amount of non-neutral surface area of the first molecular entity as at least a component of a term in a linear formula for numerically estimating said binding affinity.

6. A method of predicting binding affinity between a first molecule and a second molecule, said method comprising: determining a van der Waals interaction energy (vdW) between the first molecule and the second molecule; determining a surface contact area of the first molecule forming complimentary polar interactions with the second molecule (Att_pol); determining a surface contact area of the first molecule forming un-complimentary polar interactions with the second molecule (Rep_pol); calculating a value of pK.sub.i using at least the determined values of vdW, Att_pol, and Rep_pol.

7. The method of claim 6, wherein calculating a value of pK.sub.i comprises calculating the value of pK.sub.i using a formula pK.sub.i=C0+C1*vdW+C2*Att_pol+C3*Att_pol*Att_pol+C4*Rep_pol*Rep_pol, wherein C0, C1, C2, C3 and C4 are constant coefficients.

8. The method of claim 7, wherein C2 is greater than zero, C3 and C4 are less than zero.

9. The method of claim 6, further comprising: determining a number of rotatable bonds in the first and second molecules (Rotatable_bond); and calculating a value of pK.sub.i using a formula pK.sub.i=C0+C1*vdW+C2*Att_pol+C3*Att_pol*Att_pol+C4*Rep_pol*Rep_pol+C5*Rotatable_bond, using the determined values of vdW, Att_pol, Rep_pol, and Rotatable_bond, wherein C0, C1, C2, C3, C4 and C5 are constant coefficients.

10. The method of claim 6, wherein calculating a value of pK.sub.i comprises calculating the value of pK.sub.i using a formula pK.sub.i=C0+C1*vdW+C2*Att_pol+C3*(Att_pol*Att_pol+Rep_pol*Rep_pol), based on the determined values of vdW, Att_pol, and Rep_pol, wherein C0, C1, C2 and C3 are constant coefficients.

11. The method of claim 6, wherein determining a van der Waals interaction energy comprises determining the van der Waals energy using a grid-based approximation method.

12. The method of claim 11, wherein determining a van der Waals interaction energy comprises: representing the van der Waals interaction energy with one or more original non-linear functions; transforming each of the original non-linear functions into a moderated non-linear function, the moderated non-linear function being less non-linear than the original non-linear function; for the each of the moderated non-linear functions, applying a trilinear interpolation process to the moderated non-linear function to receive a result; reverse-transforming each of the received result; combining the reverse-transformed results; and identifying the combined result as the van der Waals interaction energy between the molecule and the **protein**.

13. A system for predicting binding affinity between a first chemical entity and a second chemical entity, said system comprising: a van der Waals energy determination module configured to determine a van der Waals interaction energy between the two chemical entities; a complimentary surface area determination module configured to define a surface area of the first chemical entity forming complimentary polar interactions with the second chemical entity; an uncomplimentary surface area determination module configured to define a surface area of the first chemical entity forming uncomplimentary polar interactions with the second chemical entity; and a calculation module configured to estimate binding affinity between the two chemical entities, using at least the van der Waals energy, the complimentary surface area, and the uncomplimentary surface area.

14. The system of claim 13, wherein the calculation module is configured to calculate a prediction value that represents the predicted binding affinity, the prediction value being at least the sum of a first coefficient, the van der Waals energy multiplied by a second coefficient, the complimentary surface area multiplied by a third coefficient, the square of the complimentary surface area multiplied by a fourth coefficient, and the square of the uncomplimentary surface area multiplied by a fifth coefficient, wherein the first, second, third, fourth and fifth coefficients are constants.

15. The system of claim 14, further comprising a coefficient determination module configured to determine respective values of the first coefficient, the second coefficient, the third coefficient, the fourth coefficient and the fifth coefficient.

16. The system of claim 14, wherein the van der Waals energy

determination module is configured to determine a van der Waals energy using a grid-based computation process, wherein the grid-based computation process comprises transforming each of one or more original non-linear functions representing the van der Waals energy into a moderated non-linear function, the moderated non-linear function being less non-linear than the original non-linear function.

17. A method of predicting binding affinity, comprising: providing a plurality of training items, each of the plurality of training items including a ligand and a **protein**; obtaining, for each of the plurality of training items, a van der Waals interaction energy between the ligand and the **protein** of the training item (vdW); obtaining, for each of the plurality of training items, a surface area of the ligand forming complimentary polar interactions with the **protein** (Att_pol); obtaining, for each of the plurality of training items, a surface area of the ligand forming un-complimentary polar interactions with the **protein** (Rep_pol); obtaining, for each of the plurality of training items, a binding affinity between the ligand and the **protein** (pK.sub.i); determining values of C0, C1, C2 and C3 using a regression technique for the formula pK.sub.i=C0+(C1*vdW)+(C2*Att_pol*Att_pol)+(C3*(Att_pol.sup.2+Rep_pol.sup.2)); and estimating an unknown binding affinity using the formula pK.sub.i=C0+C1*vdW+C2*Att_pol+C3*(Att_pol*Att_pol+Rep_pol*Rep_pol) with the determined values of C0, C1, C2 and C3.

18. A method of estimating van der Waals interaction energy comprising: transforming a function defining said van der Waals energy to a more linear functional form; computing an estimate of van der Waals energy using the more linear function; and transforming the result to correspond to the original less linear functional form.

19. The method of claim 18, wherein said computing an estimate comprises linear interpolation.

L3 ANSWER 5 OF 73 USPATFULL
AN 2002:54629 USPATFULL
PI US 2002031790 A1 20020314
PI US 2002031790 A1 20020314
TI Methods for validating **polypeptide** targets that correlate to cellular phenotypes
AB Generally applicable methods for using phenotypic probes to reduce or eliminate false positives, and thereby identify physiologically relevant endogenous target molecules, are provided. The methods use both protein interaction assay steps and phenotypic assay steps. In some embodiments, protein interactions are detected utilizing yeast two hybrid techniques.
CLM What is claimed is:
1. A method for reducing false positives from an assay that identifies **protein** interactions, comprising the steps of: a) selecting a pool of putative target molecules that interact with a first phenotypic probe in a first **protein** interaction assay; b) selecting a pool of second independent probes that interact with the pool of putative target molecules in a second **protein** interaction assay; c) selecting from the pool of second independent probes at least one confirmatory phenotypic probe that is capable of altering a phenotype of interest in a phenotypic assay host cell; and d) identifying members of the pool of putative target molecules that interact with both the first phenotypic probe and the confirmatory phenotypic probe.

2. A method for identifying a physiologically relevant target molecule that correlates to a phenotype of interest, comprising the steps of:

(a) determining a first **protein**-ligand interaction between a pool of target molecules and a first physiologically relevant probe that confers a first phenotype of interest on a host cell; (b) determining a second **protein**-ligand interaction between the pool of target molecules and a second independent physiologically relevant probe that confers a second phenotype of interest on a host cell; and (c) isolating any target molecule that interacts with both of the first and second probes.

3. The method of claim 2, wherein the first and second **protein** -ligand interactions are determined by performing a first and second yeast two-hybrid assay.

4. The method of claim 3, wherein the first yeast two-hybrid assay utilizes the pool of target molecules as prey and the second yeast two-hybrid assay uses the pool of target molecules as bait.

5. The method of claim 2, wherein said first and said second phenotypes of interest are the same cellular characteristic.

6. The method of claim 2, wherein said first and said second phenotypes of interest are related cellular characteristics.

7. A method for identifying a physiologically relevant target that correlates to a phenotype of interest, comprising the steps of: (a) exposing a primary phenotypic probe to a candidate target library; (b) identifying a pool of putative target molecules that interact with the primary phenotypic probe; (c) exposing the pool of putative target molecules to a library of candidate secondary probes; (d) identifying a sublibrary within said library of candidate secondary probes that interacts with the pool of putative target molecules; (e) selecting from said sublibrary a confirmatory probe that alters a phenotype of interest in a host cell; and (f) identifying members of the pool of putative target molecules that interact with the confirmatory probe.

8. The method of claim 7, wherein the pool of putative target molecules are perturbagen binding partners.

9. The method of claim 8, wherein said perturbagen binding partners are **polypeptides**.

10. The method of claim 7, wherein the candidate target library is an expression library of recombinant **polypeptides**.

11. The method of claim 10, wherein the expression library is encoded by genomic DNA.

12. The method of claim 10, wherein the expression library is encoded by cDNA.

13. The method of claim 7, wherein the primary and secondary phenotypic probes are perturbagens.

14. The method of claim 13, further comprising the step of fusing at least one of the perturbagens to a stabilizing **polypeptide**.

15. The method of claim 14, wherein the stabilizing **polypeptide** is GFP.

16. The method of claim 7, wherein the steps of exposing the primary and secondary probes to the pool of target molecules are performed by a first and a second yeast two-hybrid assay.

17. The method of claim 16, wherein the first yeast two-hybrid assay utilizes members of the candidate target library as prey and the second yeast two-hybrid assay uses the pool of target molecules as bait.

18. The method of claim 16, further comprising the step of eliminating bait sequences that self-activate.

19. The method of claim 16, wherein the yeast two-hybrid system utilizes a GAL4-based reporter system.

20. The method of claim 16, wherein the yeast two-hybrid system utilizes LexA-based reporter system.

21. The method of claim 19, wherein the yeast two-hybrid system utilizes a reporter vector selected from the group consisting of pVT85, pVT87, pVT88 and pVT89.

22. The method of claim 20, wherein the yeast two-hybrid system utilizes a reporter vector selected from the group consisting of pVT86 and pVT90.

23. The method of claim 19, wherein the yeast two-hybrid system utilizes a yeast strain selected from the group consisting of yVT96 and yVT97.

24. The method of claim 20, wherein the yeast two-hybrid system utilizes a yeast strain selected from the group consisting of yVT98 and yVT99.

L3 ANSWER 7 OF 73 USPATFULL
AN 2002:48277 USPATFULL
PI US 2002028468 A1 20020307
PI US 2002028468 A1 20020307
TI NMR-solve method for rapid identification of bi-ligand drug candidates
AB Methods for rapidly identifying drug candidates that can bind to an enzyme at both a common ligand site and a specificity ligand site, resulting in high affinity binding. The bi-ligand drug candidates are screened from a focused combinatorial library where the specific points of variation on a core structure are optimized. The optimal points of variation are identified by which atoms of a ligand bound to the common ligand site are identified to be proximal to the specificity ligand site. As a result, the atoms proximal to the specificity ligand site can then be used as a point for variation to generate a focused combinatorial library of high affinity drug candidates that can bind to both the common ligand site and the specificity ligand site. Different candidates in the library can then have high affinity for many related enzymes sharing a similar common ligand site.
CLM What is claimed is:
1. A method for identifying an atom of a common ligand mimic that is proximal to an interface region; wherein the enzyme can bind a common ligand (CL) or a common ligand mimic (CL mimic) at a common ligand site (CL site) and can bind a specificity ligand (SL) at an adjacent specificity ligand site (SL site); wherein an interface region is defined as the atoms of the enzyme between the CL site and SL site, and atoms of an SL if bound to the enzyme; wherein the enzyme can catalyze a reaction mechanism involving the SL and a reactive atom of the CL; and wherein a CL reactive region is defined as the reactive atom of the CL and CL atoms immediately adjacent to the reactive atom or CL atoms immediately adjacent to the SL; comprising the steps of (a) identifying an atom of the interface region, comprising the steps of (1) binding a CL to the CL site of the enzyme; (2) perturbing an atom of the CL reactive region; and (3) identifying an NMR cross-peak corresponding to an atom that is perturbed by the perturbation of the

atom of the CL reactive region, thereby identifying an atom of the interface region; then (b) identifying an atom in the CL mimic that is proximal to the interface region, comprising the steps of (1) binding a CL mimic to the CL site; (2) perturbing the interface atom identified in step (a); and (3) identifying an NMR cross-peak corresponding to an atom of the CL mimic that is perturbed by the perturbation of the interface atom, thereby identifying an atom of the CL mimic that is proximal to the interface region.

2. The method of claim 1, wherein the enzyme has a monomer molecular weight greater than 20 kD.

3. The method of claim 2, wherein the enzyme has a monomer molecular weight greater than 35 kD.

4. The method of claim 1, wherein the enzyme has a complete molecular weight greater than 50 kD.

5. The method of claim 4, wherein the enzyme has a complete molecular weight greater than 100 kD.

6. The method of claim 1, wherein the enzyme is from a human pathogen.

7. The method of claim 1, wherein the enzyme is from bacteria.

8. The method of claim 1, wherein the enzyme is a dehydrogenase.

9. The method of claim 1, wherein the enzyme is a kinase.

10. The method of claim 1, wherein the atom of the interface region in step (b)(2) is an atom of the enzyme.

11. The method of claim 1, wherein the atom of the interface region in step (b)(2) is an atom of an SL bound to the enzyme.

12. The method of claim 1, wherein the CL is a cofactor.

13. The method of claim 12, wherein the CL is SAM (S-adenosyl methionine).

14. The method of claim 12, wherein the cofactor contains a nucleotide.

15. The method of claim 14, wherein the CL is selected from the group consisting of NAD.sup.+, NADH, NADP.sup.+, NADPH, ATP and ADP.

16. The method of claim 12, wherein the CL is selected from the group consisting of farnesyl, geranyl, geranyl-geranyl and ubiquitin.

17. The method of claim 1, wherein the atom of the CL reactive region in step (a)(2) is the reactive atom of the CL.

18. The method of claim 1, wherein the atom of the reactive region in step (a)(2) is a CL atom immediately adjacent to the reactive atom.

19. The method of claim 1, wherein the atom of the reactive region in step (a)(2) is a CL atom immediately adjacent to the SL.

20. The method of claim 1, wherein a perturbing step is achieved by chemically altering an atom.

21. The method of claim 20, wherein an atom of the CL reactive region is chemically altered by replacing a hydrogen atom with a deuterium atom.

22. The method of claim 20, wherein an atom of the enzyme in the interface region is chemically altered by site-directed mutagenesis.

23. The method of claim 1, wherein a perturbing step is achieved by chemically altering an atom immediately adjacent to the perturbed atom.

24. The method of claim 23, wherein the chemical alteration is an introduction of an atom selected from the group consisting of a paramagnetic atom and a quadrupolar atom.

25. The method of claim 1, wherein a perturbing step is achieved by irradiating an atom with radio frequency energy.

26. The method of claim 1, wherein a perturbing step results in a nuclear Overhauser enhancement effect.

27. The method of claim 1, wherein a perturbing step results in an NMR cross-peak intensity or shape change.

28. The method of claim 1, wherein a perturbing step results in a relaxation effect.

29. The method of claim 1, wherein a perturbing step results in an NMR cross-peak chemical shift change.

30. The method of claim 1, wherein an NMR cross-peak is identified using a multidimensional multinuclear method, wherein the transfer of magnetization to protons is only to or from amide protons.

31. The method of claim 1, wherein an NMR cross-peak is identified using a multidimensional multinuclear method, wherein the detectable atoms are the NH protons of **protein** at an amino acid selected from the group consisting of Asn, Gin, Arg and His.

32. The method of claim 1, wherein an NMR cross-peak is identified using a multidimensional multinuclear method, wherein the detectable atoms are the methyl protons of **protein** specifically .sup.13C-.sup.1H.sub.3 labeled at an amino acid selected from the group consisting of Leu, Thr, Ile, Val, Ala and Met.

33. The method of claim 1, wherein an NMR cross-peak is identified using a multidimensional multinuclear method that includes a .sup.1H-.sup.15N correlation.

34. The method of claim 33, wherein the NMR method is a .sup.1H-.sup.5N correlation and nuclear Overhauser enhancement spectroscopy experiment.

35. The method of claim 1, wherein an NMR cross-peak is identified using a multidimensional multinuclear method that includes a .sup.1H-.sup.13C correlation.

36. The method of claim 33, wherein the NMR method is an HNCA experiment.

37. The method of claim 1, wherein an NMR cross-peak is identified using an NMR method that includes a {.sup.1H,.sup.1H} NOESY step.

38. The method of claim 37, further comprising the step of introducing a third dimension for .sup.15N or .sup.13C chemical shift.

39. The method of claim 37, wherein diagnostic .sup.1H-.sup.13C or

.sup.1H-.sup.15N one bond coupling constants are obtained by not decoupling to a heteroatom in one of the two dimensions.

40. The method of claim 37, further comprising the step of using 2D .sup.13C-.sup.1H or .sup.15N-.sup.1H HMQC or HSQC-{.sup.1H,.sup.1H} NOESY.

41. The method of claim 1, wherein an NMR cross-peak is identified using an NMR experiment that uses transverse relaxation-optimized spectroscopy (TROSY), whereby narrow line widths are achieved.

42. The method of claim 1, wherein an NMR cross-peak is identified using an NMR experiment that uses deuterium labeling and decoupling, whereby narrow line widths are achieved.

43. A method for identifying an atom of a common ligand mimic that is proximal to an interface region; wherein the enzyme can bind a common ligand (CL) or a common ligand mimic (CL mimic) at a common ligand site (CL site) and can bind a specificity ligand (SL) at an adjacent specificity ligand site (SL site); wherein an interface region is defined as the atoms of the enzyme between the CL site and SL site, and atoms of an SL if bound to the enzyme; wherein the enzyme can catalyze a reaction mechanism involving the SL and a reactive atom of the CL; wherein a CL reactive region is defined as the reactive atom of the CL and CL atoms immediately adjacent to the reactive atom or CL atoms immediately adjacent to the SL; and wherein an atom of the interface region has been identified; comprising the steps of (1) binding a CL mimic to the CL site; (2) perturbing the identified atom of the interface region; and (3) identifying an NMR cross-peak corresponding to an atom of the CL mimic that is perturbed by the perturbation of the interface atom, thereby identifying an atom of the CL mimic that is proximal to the interface region.

44. A method for identifying an atom of a common ligand mimic that is proximal to an interface region; wherein the enzyme can bind a common ligand (CL) or a common ligand mimic (CL mimic) at a common ligand site (CL site) and can bind a specificity ligand (SL) at an adjacent specificity ligand site (SL site); wherein an interface region is defined as the atoms of the enzyme between the CL site and SL site, and atoms of an SL if bound to the enzyme; wherein the enzyme can catalyze a reaction mechanism involving the SL and a reactive atom of the CL; wherein a CL reactive region is defined as the reactive atom of the CL and CL atoms immediately adjacent to the reactive atom or CL atoms immediately adjacent to the SL; comprising the steps of (1) binding a CL to the CL site in the presence of unbound CL mimic; (2) perturbing an atom of the CL, whereby energy is transferred from the CL atom to the interface region; (3) allowing the CL to unbind and a CL mimic to bind at the same CL site, whereby energy is transferred from the interface region to perturb an atom in the CL mimic; and (4) identifying an NMR cross-peak corresponding to the atom of the CL mimic perturbed in step (3), thereby identifying an atom of the CL mimic that is proximal to the interface region.

45. A method for identifying an atom of a specificity ligand mimic that is proximal to an interface region; wherein the enzyme can bind a specificity ligand (SL) or a specificity ligand mimic (SL mimic) at a specificity ligand site (SL site) and can bind a common ligand (CL) or common ligand mimic (CLM) at an adjacent common ligand site (CL site); wherein an interface region is defined as the atoms of the enzyme between the SL site and CL site, and atoms of a CL if bound to the enzyme; wherein the enzyme can catalyze a reaction mechanism involving a CL and a reactive atom of a SL; and wherein a SL reactive region is

defined as the reactive atom of the SL and SL atoms immediately adjacent to the reactive atom or SL atoms immediately adjacent to the CL; comprising the steps of (a) identifying an atom of the interface region, comprising the steps of (1) binding an SL to the SL site of the enzyme; (2) perturbing an atom of the SL reactive region; and (3) identifying an NMR cross-peak corresponding to an atom that is perturbed by the perturbation of the atom of the SL reactive region, thereby identifying an atom of the interface region; then (b) identifying an atom in the SL mimic that is proximal to the interface region, comprising the steps of (1) binding an SL mimic to the SL site; (2) perturbing the interface atom identified in step (a); and (3) identifying an NMR cross-peak corresponding to an atom of the SL mimic that is perturbed by the perturbation of the interface atom, thereby identifying an atom of the SL mimic that is proximal to the interface region.

46. A method for identifying an atom of a first ligand mimic that is proximal to an interface region; wherein the enzyme can bind a first ligand (L1) or a first ligand mimic (L1 mimic) at a first ligand site (L1 site) and can bind a second ligand (L2) at an adjacent second ligand site (L2 site); wherein an interface region is defined as the atoms of the enzyme between the L1 site and L2 site, and atoms of L2 if bound to the enzyme; wherein the enzyme can catalyze a reaction mechanism involving the L2 and L1; and wherein a L1 reactive region is defined as the reactive atom of L1, and L1 atoms immediately adjacent to the reactive atom or L1 atoms immediately adjacent to L2; comprising the steps of (a) identifying an atom of the interface region, comprising the steps of (1) binding an L1 to the L1 site of the enzyme; (2) perturbing an atom of the L1 reactive region; and (3) identifying an NMR cross-peak corresponding to an atom that is perturbed by the perturbation of the atom of the L1 reactive region, thereby identifying an atom of the interface region; then (b) identifying an atom in the L1 mimic that is proximal to the interface region, comprising the steps of (1) binding a L1 mimic to the L1 site; (2) perturbing the interface atom identified in step (a); and (3) identifying an NMR cross-peak corresponding to an atom of the L1 mimic that is perturbed by the perturbation of the interface atom, thereby identifying an atom of the L1 mimic that is proximal to the interface region.

47. A method for generating a focused combinatorial library of bi-ligand compounds that can simultaneously bind to a CL site and an SL site of an enzyme, comprising the steps of (a) performing the method of claim 1 to identify a CL mimic atom that is proximal to the interface region; and (b) synthesizing at least two compounds by modifying at least one proximal atom of the CL mimic by attaching a substituent group to the proximal atom.

48. The method of claim 47, wherein the substituent group contains a linker arm.

49. The method of claim 48, wherein the linker connects the CL mimic to a second moiety, whereby the CL mimic binds to the CL site and the second moiety binds to the SL site.

50. A combinatorial library of bi-ligand compounds obtained by the method of 49.

51. The library of claim 50, wherein the library contains at least 10 bi-ligand compounds.

52. A method for screening bi-ligand compounds, comprising the steps of (a) performing the method of claim 47 to generate a combinatorial

library of bi-ligand compounds; (b) measuring the binding of the compounds to the enzyme; and (c) identifying a compound with greater binding than the CL mimic.

53. A bi-ligand compound identified by the screening method of claim 52.

54. The bi-ligand compound of claim 53, wherein the compound reduces the activity of the enzyme.

55. The bi-ligand compound of claim 53, wherein the compound's binding affinity to the enzyme is at least 200 times greater than the CL mimic's binding affinity.

56. The bi-ligand compound of claim 55, wherein the compound's binding affinity to the enzyme is at least 1000 times greater than the CL mimic's binding affinity.

57. The bi-ligand compound of claim 56, wherein the compound's binding affinity to the enzyme is at least 5000 times greater than the CL mimic's binding affinity.

58. The bi-ligand compound of claim 53, wherein the compound's binding affinity is at least 200 times greater to the enzyme than to another enzyme in the same gene family.

L3 ANSWER 19 OF 73 USPATFULL
AN 2002:16876 USPATFULL
PI US 2002009756 A1 20020124
PI US 2002009756 A1 20020124
TI Algorithmic design of **peptides** for binding and/or modulation of the functions of receptors and/or other **proteins**
AB Methods of designing protein-targeted peptides or peptide analogues whose sequences are derived from the target protein sequences, using target protein sequence, analytically derived templates, and relevant distributions of amino acids for weighted random assignments to those templates. The templates are derived from eigenvectors of the autocovariance matrices of the physicochemically-transformed amino acid sequence of the target proteins; wavelet subsequence templates derived from wavelet transformations of the physicochemically-transformed amino acid sequence of the target proteins; and/or non-overlapping redundant subsequence templates computed from the physicochemically-transformed target protein amino acid sequence. The protein targets include cell receptors; transporters; enzymes; chaperonins; antibodies; surface proteins of infectious agents; and any protein involved in protein-protein interactions. The peptides are designed to bind to and/or otherwise modulate the function of the target protein. Partitioned amino acid distributions for weighted random assignments to the similarly partitioned templates are derived from a variety of physiologically relevant amino acid pools or regions in the target protein sequence relevant to the construction of the templates. Sequential pattern ("mode") matches between candidate peptides and their target proteins are designed such that when examined by maximum entropy, all poles power spectral transformations and/or wavelet transformations, they yield peaks of wavenumbers that differ by .ltoreq.10% of the larger wavenumber value. Also provided are examples of such mode-matched peptides, as well as methods for their use in elucidating sites on proteins for drug design and testing, detection of disease conditions or contaminants, and as therapeutics for protein function modulation in disease treatment.
CLM What is claimed is:
1. A method for synthesizing a **peptide** based on matching a

physicochemical mode of a **peptide** to the same physicochemical mode of a target **polypeptide** or **protein**, followed by synthesizing a retro-inverso **peptide** version of said **peptide** comprised of D-amino acids, comprising the steps of: assigning a numerical value of an orderable physicochemical property to each member of a set of **peptide** constituents, said set of **peptide** constituents including all the members of the set of naturally-occurring L-amino acids; arranging said **peptide** constituents in order of said numerical values of said orderable physicochemical property; partitioning said set of **peptide** constituents into a plurality of **peptide** constituent groups, whereby each of said **peptide** constituent groups contains at least one member of said set of **peptide** constituents, each **peptide** constituent group encompasses a range of said ordered numerical values, and each member of said set of **peptide** constituents belongs to only one **peptide** constitutent group; creating a **polypeptide** physicochemical data series by replacing each amino acid in an amino acid sequence of said target **polypeptide** or **protein** with said numerical value of said orderable physicochemical property corresponding to said each amino acid in said amino acid sequence of said target **polypeptide** or **protein**; calculating one or more **polypeptide** eigenvalues and a corresponding **polypeptide** eigenvector associated with each of said one or more **polypeptide** eigenvalues by linear decomposition of an autocovariance matrix formed from a sequentially lagged data matrix of said **polypeptide** physicochemical data series; ordering said one or more **polypeptide** eigenvalues and said corresponding **polypeptide** eigenvectors from largest to smallest; selecting one or more of said **polypeptide** eigenvectors; transforming said one or more of said **polypeptide** eigenvectors into an eigenvector template; forming a graph of said eigenvector template, wherein said numerical values of said physicochemical property are graphed along the y-axis of said graph and ordered position in said eigenvector template is graphed along the x-axis of said graph; partitioning said graph along said y-axis according to said ranges of said numerical values of said physicochemical property defining said **peptide** constituent groups, to form a plurality of y-axis ranges; assigning one of said **peptide** constituents to each position in said **peptide** by using said graph as a template to create a sequence of a mode-matched **peptide**, wherein at each ordered position in said eigenvector template along said x-axis of said graph, said one of said **peptide** constituents assigned to said ordered position has a value of said orderable physicochemical property that is within said y-axis range of said ordered point; determining a sequence of a retro-inverso **peptide** by inverting said sequence of a mode-matched **peptide**; and synthesizing said retro-inverso **peptide** from said sequence, using D-amino acids.

2. A method for synthesizing a **peptide** based on matching a physicochemical mode of a **peptide** to the same physicochemical mode of a target **polypeptide** or **protein**, followed by synthesizing a retro-inverso version of said **peptide** comprised of D-amino acids, comprising the steps of: assigning a numerical value of an orderable physicochemical property to each member of a set of **peptide** constituents, said set of **peptide** constituents including all the members of the set of naturally-occurring amino acids; arranging said **peptide** constituents in order of said numerical values of said orderable physicochemical property; partitioning said set of **peptide** constituents into a plurality of **peptide** constituent groups, whereby each of said **peptide** constituent groups contains at least one member of said set of

peptide constituents, each **peptide** constituent group encompasses a range of said ordered numerical values, and each member of said set of **peptide** constituents belongs to only one **peptide** constituent group; creating a **polypeptide** physicochemical data series by replacing each amino acid in an amino acid sequence with said numerical value of said orderable physicochemical property corresponding to said each amino acid in said amino acid sequence; calculating one or more **polypeptide** eigenvalues and a corresponding **polypeptide** eigenvector associated with each of said one or more **polypeptide** eigenvalues by linear decomposition of an autocovariance matrix formed from a sequentially lagged data matrix of said **polypeptide** physicochemical data series; ordering said one or more **polypeptide** eigenvalues and said corresponding **polypeptide** eigenvectors from largest to smallest; selecting one or more of said **polypeptide** eigenvectors; forming a vector, said vector being a sum of the products of each of said plurality of said **polypeptide** eigenvectors multiplied by the corresponding eigenvalue; forming a graph of said vector, wherein said numerical values of said orderable physicochemical property are graphed along the y-axis of said graph, and ordered position in said eigenvector template is graphed along the x-axis of said graph; partitioning said graph along said y-axis according to said range of said numerical values of said orderable physicochemical property defining said **peptide** constituent groups, to form a plurality of y-axis ranges; and assigning one of said **peptide** constituents to each position in said **peptide** by using said graph of said vector as a template, wherein at each ordered position in said eigenvector template along said x-axis of said graph, said one of said **peptide** constituents assigned to said ordered position has a value of said orderable physicochemical property that is within said y-axis range of said ordered position; determining a sequence of a retro-inverso **peptide** by inverting said sequence of a mode-matched **peptide**; and synthesizing said retro-inverso **peptide** from said sequence, using D-amino acids.

L3 ANSWER 20 OF 73 USPATFULL
AN 2002:14526 USPATFULL
PI US 6341256 B1 20020122
PI US 6341256 B1 20020122
TI Consensus configurational bias Monte Carlo method and system for pharmacophore structure determination
AB In a specific embodiment, this invention includes a method for determining an accurate, consensus pharmacophore structure shared by compounds that bind selectively to a target molecule. Optionally, the method begins with screening a diversity library against the target molecule of interest to pick the selectively binding members. Next the structure of the selected members is examined and a candidate pharmacophore responsible for the binding to the target molecule is determined. Next, preferably by REDOR nuclear magnetic resonance, several highly accurate interatomic distances are determined in certain of the selected members which are related to the candidate pharmacophore. A highly accurate consensus, configurational bias, Monte Carlo method determination of the structure of the candidate pharmacophore is made using the structure of the selected members and incorporating as constraints the shared candidate pharmacophore and the several measured distances. This determination is adapted to efficiently examine only relatively low energy configurations while respecting any structural constraints present in the organic diversity library. If the diversity library contains short peptides, the determination respects the known degrees of freedom of peptides as well as any internal

constraints, such as those imposed by disulfide bridges. Finally, the highly accurate pharmacophore so determined is used to select lead organics for drug development targeted at the initial target molecule.

CLM What is claimed is:

1. A method of determining a consensus pharmacophore for binding to a target molecule at a temperature of interest, said method comprising determining a consensus structure for each of one or more **peptides** or **peptide** derivatives having a backbone represented by rigid molecular subunits connected by bonds, the bonds allowing torsional rotation of the rigid subunit, wherein said **peptides** or **peptide** derivatives bind to said target molecule at said temperature of interest, and wherein said determining of a consensus structure employs a consensus configurational bias Monte Carlo algorithm and comprises steps of repeatedly: (a) generating proposed structures for each one of said **peptides** or **peptide** derivatives according to moves comprising constrained concerted torsional rotations about a limited region of the backbone by a method comprising (i) making a torsional angle rotation about a chosen backbone bond, and (ii) choosing subsequent backbone torsional rotations so that at least one and at most four contiguous rigid subunits of the backbone undergo a spatial displacement; and (b) accepting a proposed structure for each one of said **peptides** or **peptide** derivatives according to a configurationally-biased Metropolis acceptance probability depending on a molecular Hamiltonian further including one or more heuristic constraint terms determined from the proposed structures for each one of said **peptides** or **peptide** derivatives, until sufficient structures have been generated and accepted to permit a statistically significant determination of a consensus structure for each **peptide** or **peptide** derivative, and wherein said consensus pharmacophore comprises interatomic distances between selected chemical groups in each of said consensus structures.

2. The method of claim 1, wherein said one or more **peptides** or **peptide** derivatives are identified by screening one or more diversity libraries for a plurality of **peptides** or **peptide** derivatives that bind to said target molecule, said screening comprising contacting said target molecule with **peptides** or **peptide** derivatives in said diversity libraries.

3. The method of claim 2, wherein said screening further comprises using a genetic selection technique.

4. The method of claim 1 or 2, further comprising the step of measuring one or more interatomic distances in one or more of said **peptides** or **peptide** derivatives.

5. The method of claim 4, wherein the step of measuring one or more interatomic distances comprises a step of making solid phase nuclear magnetic resonance measurements on selected nuclei in a sample comprising one of the **peptides** or **peptide** derivatives.

6. The method of claim 4, wherein the step of measuring one or more interatomic distances comprises a step of making liquid phase nuclear magnetic resonance measurements on selected nuclei in a sample comprising one of the binding compounds.

7. The method of claim 5, wherein the selected nuclei are selected from the group consisting of .sup.13C, .sup.15N, .sup.19F, and .sup.31P.

8. The method of claim 5, wherein said one of the **peptides** or **peptide** derivatives is bound to the target molecule.

9. The method of claim 5, wherein said one of the **peptides** or **peptide** derivatives of said sample is covalently attached to a surface of a substrate during said step of making solid phase nuclear magnetic resonance measurements.

10. The method of claim 5, wherein the solid phase nuclear magnetic resonance measurements are made by means of REDOR NMR.

11. The method of claim 9, wherein a plurality of molecules of said one of the **peptides** or **peptide** derivatives is individually attached to the substrate at a surface density such that the inter-nuclear dipole-dipole interactions between different molecules is less than 10% of inter-nuclear dipole-dipole interactions within one molecule.

12. The method of claim 11, wherein said plurality of molecules of said one of the **peptides** or **peptide** derivatives is at least 95% pure.

13. The method of claim 9, wherein the substrate has pores of sufficient size to permit a molecule of the target compound to diffuse and bind to a molecule of said one of the **peptides** or **peptide** derivatives.

14. The method of claim 9, wherein the substrate is selected from the group consisting of p-MethylBenzhydrilamine resin, divinylbenzyl polystyrene resin, and glass beads.

15. The method of claim 1 or 2 wherein the **peptides** or **peptide** derivatives are constrained by internal bonds.

16. The method of claim 15, wherein the internal bonds are disulfide bonds.

17. The method of claim 15 wherein the **peptides** or **peptide** derivatives contain pairs of cysteine residues.

18. The method of claim 17, wherein the cysteine residues are separated by 2 to 16 amino acid residues.

19. The method of claim 18, wherein the cysteine residues are separated by 6 to 8 amino acid residues.

20. The method of claim 1 or 2 wherein a consensus structure is determined for each of one or more **peptides** having the formula R.sup.1CX.sub.nCR.sup.2, wherein: R.sup.1 is a first sequence of 0 to 10 amino acid residues; R.sup.2 is a second sequence of 0 to 10 amino acid residues; X.sub.n is a sequence of n amino acid residues; and n is an integer ranging from 2 to 16.

21. The method of claim 1, wherein the energy is determined from the proposed structures for each of said **peptides** or **peptide** derivatives by means of a Hamiltonian which includes constraint terms which represent distance measurements made for each of said **peptides** or **peptide** derivatives.

22. The method of claim 21, wherein the constraint terms comprise a weighted sum of squares of differences of interatomic distances of the proposed structure and measured interatomic distances.

23. The method of claim 1, wherein said determining of a consensus structure for each of the one or more **peptides** or **peptide** derivatives further comprises after said steps of repeatedly generating and accepting proposed structures, steps of: (a) clustering the accepted proposed structures for each one of said **peptides** or **peptide** derivatives, and (b) averaging the accepted proposed structures for each one of said **peptides** or **peptide** derivatives in each cluster; wherein the average proposed structure for a particular **peptide** or **peptide** derivative in a particular cluster is a consensus structure for the particular **peptide** or **peptide** derivative in the particular cluster, and wherein the consensus pharmacophore comprises interatomic distances between selected groups in the consensus structures of the **peptides** or **peptide** derivatives in each cluster.

24. A method for determining a consensus pharmacophore for binding to a target molecule comprising: (a) screening one or more diversity libraries to select a plurality of **peptides** or **peptide** derivatives that bind to said target molecule, wherein said screening comprises contacting said target molecule with compounds in said diversity library and wherein said **peptides** or **peptide** derivatives have conformational degrees of freedom at a temperature of interest limited to torsional rotations of rigid molecular subunits about bonds between said subunits; and (b) determining a consensus structure of each one of said **peptides** or **peptide** derivatives by a method employing a consensus configurational bias Monte Carlo algorithm and comprises steps of repeatedly (i) generating proposed structures for each one of said **peptides** or **peptide** derivatives according to moves comprising constrained concerted torsional rotations about a limited region of a backbone represented by rigid molecular subunits connected by bonds, the bonds allowing torsional rotation of the rigid subunits by a method comprising (A) making a torsional angle rotation about a chosen backbone bond, and (B) choosing subsequent backbone torsional rotations so that at least one and at most four contiguous rigid subunits of the backbone undergo a spatial displacement; (ii) accepting a proposed structure for each one of said **peptides** or **peptide** derivatives according to a configurationally-biased Metropolis acceptance probability depending on a molecular Hamiltonian further including one or more heuristic constraint terms determined from the proposed structures for each one of said **peptides** or **peptide** derivatives until sufficient structures have been generated and accepted to permit a statistically significant determination of a consensus structure for each one of said **peptides** or **peptide** derivatives, wherein said consensus pharmacophore comprises interatomic distances between selected chemical groups in each of said consensus structures.

25. The method of claim 24 wherein the compounds that bind to said target molecule are **peptides** comprising the formula R.sup.1CX.sub.nCR.sup.2, wherein: R.sup.1 is a first sequence of 0 to 10 amino acid residues; R.sup.2 is a second sequence of 0 to 10 amino acid residues; X.sub.n is a sequence of n amino acid residues; and n is an integer ranging from 2 to 16.

26. A method of determining a consensus pharmacophore for binding to a target molecule comprising: (a) screening one or more diversity libraries to select a plurality of **peptides** or **peptide** derivatives that bind to said target molecule, wherein said screening comprises contacting said target molecule with compounds in said diversity library and wherein said **peptides** or **peptide**

derivatives have a backbone represented by rigid molecular subunits connected by bonds, the bonds allowing torsional rotation of the rigid subunits; (b) measuring one or more interatomic distances in one or more of said **peptides** or **peptide** derivatives; and (c) determining a consensus structure for each of said **peptides** or **peptide** derivatives by means of a consensus configurational bias Monte Carlo algorithm comprising steps of repeatedly (i) generating proposed structures for each one of said **peptides** or **peptide** derivatives according to moves comprising constrained concerted torsional rotations in a limited region of the backbone by a method comprising (A) making a torsional angle rotation about a chosen backbone bond, and (B) choosing subsequent backbone torsional rotations so that at least one and at most four contiguous rigid subunits of the backbone undergo a spatial displacement; (ii) accepting a proposed structure for each one of said **peptides** or **peptide** derivatives according to a configurationally-biased Metropolis acceptance probability depending on a molecular Hamiltonian further including one or more heuristic constraint terms determined from the proposed structures for each one of said **peptides** or **peptide** derivatives until sufficient structures have been generated and accepted to permit a statistically significant determination of a consensus structure for each **peptide** or **peptide** derivative, wherein said consensus pharmacophore comprises interatomic distances between selected chemical groups in each of said consensus structures.

27. The method of claim 26, wherein the step of measuring one or more interatomic distances comprises a step of making solid phase nuclear magnetic resonance measurements on selected nuclei in a sample comprising one of the **peptides** or **peptide** derivatives.

28. The method of claim 27, wherein the solid phase nuclear magnetic resonanance measurements are made by means of REDOR NMR.

29. A method of determining a consensus pharmacophore for binding to a target molecule at a temperature of interest, said method comprising: (a) providing a library comprising one or more **peptides** or **peptide** derivatives, wherein each **peptide** or **peptide** derivative binds to the target at the temperature of interest, each member of the library comprising a candidate pharmacophore; (b) obtaining structural constraints for at least one of the **peptides** or **peptide** derivatives in the library; c) determining a structure for each of the one or more **peptides** or **peptide** derivatives wherein each **peptide** or **peptide** derivative is represented as having a backbone comprising rigid molecular subunits and wherein said subunits are related by torsional rotations about bonds between said subunits, and wherein said determining of the structure comprises use of a consensus configurational bias Monte Carlo algorithm, the algorithm comprising steps of: (i) generating constrained concerted torsional rotations about a limited region of the backbone by a method comprising (A) making a torsional angle rotation about a chosen backbone bond, and (B) choosing subsequent backbone torsional rotations so that no more than four rigid subunits of the backbone undergo a spatial displacement; and (ii) determining whether to accept the structure so obtained for each one of said **peptides** or **peptide** derivatives according to a configurationally-biased Metropolis acceptance probability depending on a molecular Hamiltonian further including one or more heuristic constraint terms determined from the proposed structures for each one of said **peptides** or **peptide** derivatives; and d) including an accepted structure for a **peptide** or

peptide derivative as a shared structure, until sufficient shared structures have been accepted to permit a statistically significant determination of a consensus structure for the library of **peptides** or **peptide** derivatives, wherein said consensus structure comprises the consensus pharmacophore and wherein the structure of said consensus pharmacophore comprises selected chemical groups consistent with the structural constraints obtained in step (b).

30. The method of claim 29 wherein obtaining the structural constraints of step (b) comprises the step of measuring one or more interatomic distances in one or more of said **peptides** or **peptide** derivatives.

31. The method of claim 30, wherein the step of measuring one or more interatomic distances comprises a step of making solid phase nuclear magnetic resonance measurements on selected nuclei in a sample comprising one of the **peptides** or **peptide** derivatives.

32. The method of claim 30, wherein the step of measuring one or more interatomic distances comprises a step of making liquid phase nuclear magnetic resonance measurements on selected nuclei in a sample comprising one of the binding compounds.

33. The method of claim 31, wherein the selected nuclei are selected from the group consisting of .sup.13C, .sup.15N, .sup.19F, and .sup.31P.

34. The method of claim 31, wherein said one of the **peptides** or **peptide** derivatives of said sample is covalently attached to a surface of a substrate during said step of making solid phase nuclear magnetic resonance measurements.

35. The method of claim 31, wherein the solid phase nuclear magnetic resonance measurements are made by means of REDOR NMR.

36. The method of claim 34, wherein a plurality of molecules of said one of the **peptides** or **peptide** derivatives is individually attached to the substrate at a surface density such that the inter-nuclear dipole-dipole interactions between different molecules is less than 10% of inter-nuclear dipole-dipole interactions within one molecule.

37. The method of claim 34, wherein the substrate has pores of sufficient size to permit a molecule of the target compound to diffuse and bind to a molecule of said one of the **peptides** or **peptide** derivatives.

38. The method of claim 34, wherein the substrate is selected from the group consisting of p-MethylBenzhydrilamine resin, divinylbenzyl polystyrene resin, and glass beads.

39. The method of claim 36, wherein said plurality of molecules of said one of the **peptides** or **peptide** derivatives is at least 95% pure.

40. The method of claim 29 wherein said one or more **peptides** or **peptide** derivatives are identified by screening one or more diversity libraries for a plurality of **peptides** or **peptide** derivatives that bind to said target molecule, said screening comprising contacting said target molecule with **peptides** or **peptide** derivatives in said diversity

libraries.

41. The method of claim 29 wherein the **peptides** or **peptide** derivatives are constrained by internal bonds.

42. The method of claim 41 wherein the **peptides** or **peptide** derivatives contain pairs of cysteine residues.

43. The method of claim 42 wherein the cysteine residues are separated by 2 to 16 amino acid residues.

44. The method of claim 43, wherein the cysteine residues are separated by 6 to 8 amino acid residues.

45. The method of claim 41 wherein the internal bonds are disulfide bonds.

46. The method of claim 29 wherein a consensus structure is determined for each of one or more **peptides** having the formula R.sup.1CX.sub.nCR.sup.2, wherein: R.sup.1 is a first sequence of 0 to 10 amino acid residues; R.sup.2 is a second sequence of 0 to 10 amino acid residues; X.sub.n is a sequence of n amino acid residues; and n is an integer ranging from 2 to 16.

47. The method of claim 29 wherein said determining of a consensus structure for each of the one or more **peptides** or **peptide** derivatives further comprises, after said steps of repeatedly generating and accepting proposed structures, steps of: (a) clustering the accepted proposed structures for each one of said **peptides** or **peptide** derivatives, and (b) averaging the accepted proposed structure for each one of said **peptides** or **peptide** derivatives in each cluster; wherein the average proposed structure for a particular **peptide** or **peptide** derivative in a particular cluster is a consensus structure for the particular **peptide** or **peptide** derivative in the particular cluster, and wherein the consensus pharmacophore comprises interatomic distances between selected groups in the consensus structures of the **peptides** or **peptide** derivatives in each cluster.

48. The method of claim 29 wherein the heuristic terms of the energy Hamiltonian includes a measurement constraint term.

49. The method of claim 48 wherein the measurement constraint term for a given atom pair in the **peptide** or **peptide** derivative is given by: ##EQU24## where: R.sub.l,ij is the distance between the i-th and j-th atom in the l-th **peptide** or **peptide** derivative; R.sup.(o).sub.l,ij is the measured distance between the i-th and j-th atom in the l-th **peptide** or **peptide** derivative; and w.sub.l,ij is a weighting factor for the i-j atom pair in the l-th **peptide** or **peptide** derivative.

50. The method of claim 49 wherein the weighting factor depends, in part, on the measured distance of the atom pair.

51. The method of claim 50 wherein the weighting factor favors measured pair distances that are less than 3 .ANG..

52. The method of claim 50 wherein the weighting factor disregards measured pair distances that are greater than 7 .ANG..

53. The method of claim 29 wherein the heuristic terms of the energy

Hamiltonian includes a consensus constraint term.

54. The method of claim 53 wherein the consensus constraint term for an atom pair in a **peptide** or **peptide** derivative is given by: ##EQU25## where: R.sub.l,ij is the distance between the i-th and j-th atom in the l-th **peptide** or **peptide** derivative; R.sup.(c).sub.ij is the averaged distance between the i-th and j-th atom averaged over all **peptides** or **peptide** derivatives; and w'.sub.l,ij is a weighting factor for the i-j atom pair in the l-th **peptide** or **peptide** derivative.

55. The method of claim 54 wherein the weighting factor depends, in part, on the known affinity of the **peptide** or **peptide** derivative to the target molecule.

56. The method of claim 55 wherein the weighting factor is proportional to the logarithm of the known affinity of the binder to the target molecule.

L3 ANSWER 22 OF 73 USPATFULL
AN 2002:8706 USPATFULL
PI US 2002004706 A1 20020110
PI US 2002004706 A1 20020110
TI Apparatus and method for automated **protein** design
AB The present invention relates to apparatus and methods for quantitative protein design and optimization.
CLM What is claimed is:

1. A method executed by a computer under the control of a program, said computer including a memory for storing said program, said method comprising the steps of: (A) receiving a **protein** backbone structure with variable residue positions; (B) establishing a group of potential rotamers for each of said variable residue positions, wherein at least one variable residue position has rotamers from at least two different amino acid side chains; and (C) analyzing the interaction of each of said rotamers with all or part of the remainder of said **protein** backbone structure to generate a set of optimized **protein** sequences, wherein said analyzing step includes a Dead-End Elimination (DEE) computation.

2. A method executed by a computer under the control of a program, said computer including a memory for storing said program, said method comprising the steps of: (A) receiving a **protein** backbone structure with variable residue positions; (B) classifying each variable residue position as either a core, surface or boundary residue; (C) establishing a group of potential rotamers for each of said variable residue positions, wherein at least one variable residue position has rotamers from at least two different amino acid side chains; and (D) analyzing the interaction of each of said rotamers with all or part of the remainder of said **protein** to generate a set of optimized **protein** sequences.

3. A method according to claim 2 wherein said analyzing step comprises a DEE computation.

4. A method according to claim 1 or 2 wherein said set of optimized **protein** sequences comprises the globally optimal **protein** sequence.

5. A method according to claim 1 or 3 wherein said DEE computation is selected from the group consisting of original DEE and Goldstein DEE.

6. A method according to claim 1 or 2 wherein said analyzing step includes the use of at least one scoring function.

7. A method according to claim 6 wherein said scoring function is selected from the group consisting of a Van der Waals potential scoring function, a hydrogen bond potential scoring function, an atomic solvation scoring function, an electrostatic scoring function and a secondary structure propensity scoring function.

8. A method according to claim 6 wherein said analyzing step includes the use of at least two scoring functions.

9. A method according to claim 6 wherein said analyzing step includes the use of at least three scoring functions.

10. A method according to claim 6 wherein said analyzing step includes the use of at least four scoring functions.

11. A method according to claim 1 or 2 further comprising testing at least one member of said set to produce experimental results.

12. A method according to claim 4 further comprising (D) generating a rank ordered list of additional optimal sequences from said globally optimal **protein** sequence.

13. A method according to claim 12 wherein said generating includes the use of a Monte Carlo search.

14. A method according to claim 2 wherein said analyzing step step comprises a Monte Carlo computation.

15. A method according to claim 12 further comprising: (E) testing some or all of said **protein** sequences from said ordered list to produce potential energy test results.

16. A method according to claim 15 further comprising: (F) analyzing the correspondence between said potential energy test results and theoretical potential energy data.

17. An optimized **protein** sequence generated by the method of claim 1 or 2.

18. A nucleic acid sequence encoding a **protein** sequence according to claim 17.

19. An expression vector comprising the nucleic acid of claim 18.

20. A host cell comprising the nucleic acid of claim 18.

21. A **protein** having a sequence that is at least about 5% different from a known **protein** sequence and is at least 20% more stable than the known **protein** sequence.

22. A computer readable memory to direct a computer to function in a specified manner, comprising: a side chain module to correlate a group of potential rotamers for residue positions of a **protein** backbone model; a ranking module to analyze the interaction of each of said rotamers with all or part of the remainder of said **protein** to generate a set of optimized **protein** sequences.

23. A computer readable memory according to claim 22 wherein said ranking module includes a van der Waals scoring function component.

24. A computer readable memory according to claim 22 wherein said ranking module includes an atomic solvation scoring function component.

25. A computer readable memory according to claim 22 wherein said ranking module includes a hydrogen bond scoring function component.

26. A computer readable memory according to claim 22 wherein said ranking module includes a secondary structure scoring function component.

27. A computer readable memory according to claim 22 further comprising an assessment module to assess the correspondence between potential energy test results and theoretical potential energy data.

L3 ANSWER 26 OF 73 USPATFULL
AN 2001:224368 USPATFULL
PI US 2001049585 A1 20011206
PI US 2001049585 A1 20011206
TI Computer predictions of molecules
AB A method for predicting a set of chemical, physical or biological features related to chemical substances or related to interactions of chemical substances including using at least 16 different individual prediction means, thereby providing an individual prediction of the set of features for each of the individual prediction means and predicting the set of features on the basis of combining the individual predictions, the combining being performed in such a manner that the combined prediction is more accurate on a test set than substantially any of the predictions of the individual prediction means.
CLM What is claimed is:
1. A method for predicting a set of chemical, physical or biological features related to chemical substances or related to interactions of chemical substances using a system comprising a plurality of prediction means, the method comprising using at least 16 different individual prediction means, thereby providing an individual prediction of the set of features for each of the individual prediction means and predicting the set of features on the basis of combining the individual predictions, the combining being performed in such a manner that the combined prediction is more accurate on a test set than substantially any of the predictions of the individual prediction means.

2. A method according to claim 1, wherein the combining being performed is an averaging and/or weighted averaging process.

3. A method according to claim 1, wherein the combining of the predictions provided by the individual prediction means are based on predictions provided by either substantially all or all prediction means of the system or substantially all or all prediction means of the system which do not compromise the accuracy of the combined prediction or substantially all or all prediction means of the system which are accurate above a given value or substantially all or all prediction means of the system which are estimated to be accurate above a given confidence rating.

4. A method according to claim 1, wherein the number of different predictions means is at least 20, such as at least 30, such as at least 40, 50, 75, 100, 200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000, 2500, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10,000, 15,000, 20,000, 30,000, 40,000, 50,000, 100,000, 200,000, 500,000, 1,000,000.

5. A method according to claim 1, wherein the type of prediction means

are selected from the group consisting of neural networks, hidden Markov models (HMM), EM algorithms, weight matrices, decision trees, fuzzy logic, dynamical programming, nearest neighbour approaches, and vector support machines.

6. A method according to claim 1, wherein the prediction means are diverse with respect to type, and/or with respect to architecture, and/or in case of prediction means subjected to training with respect to initial conditions, and/or with respect to training.

7. A method according to claim 2, wherein the weighted averaging process is performed based on the accuracy of substantially each or each of the individual prediction means.

8. A method according to claim 7, wherein the individual predictions performed are a series of predictions, and the weighting comprises an evaluation of the relative accuracy of substantially each individual prediction or each individual prediction means on substantially all, or one or more subsets of the predictions in a series of predictions.

9. A method according to claim 8, wherein the weighting of particular individual predictions means results in an evaluation the predictions rendered by the systemson substantially all or one or more of the subsets of the predictions in a series of predictions are to be excluded from the weighted average, and the individual prediction means in question is/are excluded from the weighted average in further predictions, either with respect to substantially all or with respect to one or more of the subsets of the predictions in a series of predictions.

10. A method according to claim 3, wherein the confidence rating is calculated by multiplying each component of an individual prediction of the selected prediction means by the weight obtained for a sequence and prediction means, the resulting product summed for each component of each residue over all prediction means, the resulting sums being divided by the sum of weights, and the resulting maximal per-residue component quotient being used to determine the H or E or C secondary structure assignment for that residue.

11. A method according to claim 9, wherein the number of prediction means not excluded being at least 3 such as 4, preferably at least 5, 6, 7, 8, 9, or 10.

12. A method according to claim 10, wherein the number of prediction means not excluded being at least 3 such as 4, preferably at least 5, 6, 7, 8, 9, or 10.

13. A method for establishing a prediction system for predicting a set of chemical, physical or biological features related to chemical substances or to chemical interactions represented by an input data using a system comprising a plurality of prediction means, the method comprises performing the steps according to claim 1.

14. A method according to claim 1, wherein the prediction means comprise neural networks.

15. A method according to claim 14, wherein the neural networks are different with respect to architecture, and/or with respect to initial conditions, and/or with respect to selection of training set, and/or with respect to learning rate and/or with respect to subtypes of input data fed to respective neural networks, and/or with respect to subtypes of output data sets rendered by the respective neural networks.

16. A method according to claim 1, wherein the chemical, physical or biological features related to chemical substances or to chemical interactions to be predicted are descriptors of molecules or subsets of molecules.

17. A method according to claim 16, wherein descriptors are selected from the group comprising secondary structure class assignment, tertiary structure, interatomic distance, bond strength, bond angle, descriptors relating to or reflecting hydrophobicity, hydrophilicity, acidity, basicity, relative nucleophilicity, relative electrophilicity, electron density or rotational freedom, scalar products of atomic vectors, cross products of atomic vectors, angles between atomic vectors, triple scalar products between atomic vectors, torsion angles, atomic angles such as but not exclusively omega, psi, phi, chi1, chi2, chi3, chi4, chi5 angles, chain curvature, chain torsion angles, and mathematical functions thereof.

18. A method according claim 16, wherein molecules are selected from the group comprising **proteins**, **polypeptides**, oligopeptides, **protein** analogues, peptidomimietic, **peptide** isosteres, pseudopeptide, nucleotides and derivatives thereof, PNA and nucleic acids.

19. A method according claim 18, wherein molecules are selected from the group comprising **proteins**, **peptides**, **polypeptides** and oligopeptides.

20. A method according to claim 1, wherein the prediction means of the system are arranged in levels and wherein at least one subtype of data provided by a first level of prediction means is transferred changed or unchanged to at least one subsequent level.

21. A method according to claim 20, wherein the at least one subtype of data transferred to the at least one subsequent level comprises subsets of predictions provded by the first level of prediction means and/or subtypes of input data either changed or unchanged from input data fed into the first neural network system.

22. A method according to claim 20, wherein subtypes of input data are selected from the group comprising amino acid sequence, nucleic acid sequence, sequence profile, amino acid composition, nucleic acid composition, window, window size, length of **protein**, length of nucleotide, and descriptor.

23. A method according to claim 13, wherein input data comprises input elements each having a corresponding output element, and the input elements may be arranged in one or more sequences, such as an amino acid residue or a nucleotide residue in a **peptide** or nucleic acid sequence, and that for each input element, predictions are made for more than one output element.

24. A method according to claim 23, wherein the more than one output elements correspond to neighbouring input elements.

25. A method for prediction of descriptors of **protein** structures or substructures comprising feeding input data representing at least one residue of a **protein** sequence to at least 16 diverse neural networks arranged in parallel in a first level, generating by use of the networks arranged in the first level a single- or a multi-component output for each networks the single- or multi-component output representing a descriptor of one residue

comprised in the **protein** sequence represented in the input data, or the single- or multi-component output representing a descriptor of 2 or more consecutive residues of the **protein** sequence, providing the single- or multi-component output from each network of the first level as input to one or more neural networks arranged in parallel to a subsequent level(s) in a hierarchical arrangement of levels, optionally inputting one or more subsets of the **protein** sequence and/or substantially all of the **protein** sequence to the second or subsequent level(s), generating by use of the networks arranged in the subsequent level(s) single or multi-component output data representing a descriptor for each residue in the input sequence, weighting the output data of each neural network of the subsequent level(s) to generate a weighted average for each component of the descriptor, optionally selecting from the multi-component output data, if generated, the component of descriptor with the highest weighted average as the predicted descriptor for each amino acid in the **protein** sequence, or optionally assigning a descriptor to a single-component output, and optionally assigning the descriptor of said **protein** sequence.

26. A method according to claim 25, wherein the number of neural networks in one level is at least 20, such as at least 30, such as at least 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 1000, 10000, 100 000 and 1 000 000.

27. A method according to claim 25, wherein the said neural networks are trained by a training process comprising an X-fold cross-validation procedure wherein each network was trained on (X-1) of X subsets of data and tested on 1 or more of said subsets.

28. A method according to claim 25, wherein the neural networks are trained by a training process comprising an 10-fold cross-validation procedure wherein each network was trained 9 of said subsets of data and tested on 1 of said subsets.

29. A method according to claim 25, wherein the neural networks are trained by a training process comprising supplying input data, filtered or unfiltered from a database, generating by use of the networks arranged in the first level a single- or a multi-component output for each networks, the single- or multi-component output represents a descriptor of one residue comprised in the **protein** sequence represented in the input data, or the single- or multi-component output represents a descriptor of 2 or more, consecutive residues of a **protein** sequence, providing the single- or multi-component output from each network of the first level as input to one or more neural networks arranged in parallel in a subsequent level(s) in a hierarchical arrangement of levels, optionally inputting one or more subsets of the **protein** sequence and/or substantially all of the **protein** sequence to the subsequent level(s), generating by use of the networks arranged in the second or subsequent level(s) a single or multi-component output representing a descriptor for each residue in the input sequence, weighting the output of each neural network of the subsequent level(s) to generate a weighted average for each component of the descriptor, and performing an X-fold cross-validation procedure wherein each network was trained on (X-1) of X subsets of data and tested on 1 or more subsets of data

30. A method according to claim 27, wherein X is from 2 to 1 000 0000, such as from 2 to 100 000, 2 to 10 000, 2 to 1000, 2 to 100, 2 to 50, preferably 5 to 50, such as 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50.

31. A method according to claim 27 wherein the testing on the subset

comprises making a prediction for each element in the data set and evaluating the accuracy of the prediction.

32. A method according to claim 25, wherein the one or more neural networks arranged in parallel to a subsequent level(s) in a hierarchical arrangement of levels comprises networks with at least two different window sizes, such at least 3, 4, 5, or 6 window sizes.

33. A method according to claim 25, wherein the one or more neural networks arranged in parallel to a subsequent level(s) in a hierarchical arrangement of levels comprises networks with at least 1 hidden unit, such as at least 2, 5, 10, 20, 30, 40, 50, 60, 75 or 100 hidden units.

34. A method according to claim 25, wherein the one or more neural networks arranged in parallel to a subsequent level(s) in a hierarchical arrangement of levels comprises networks with at least 7, such as at least 9, such as at least 11, particularly at least an 101 residue input window, such as at least 13, 15, 17, 21, 31, 41, 51, or 101 residue input window.

35. A method according to claim 25, wherein the single- or multi-component output from at least one neural networks in at least one level in a hierarchical arrangement of levels of neural networks is supplied as input to more than one neural network in a subsequent level of neural networks.

36. A method according to claim 25, wherein diverse networks are diverse with respect to architecture and/or initial conditions and/or selection of learning set, and/or position-specific learning rate, and/or subtypes of input data presented to respective neural networks, and or with respect to subtypes of output data sets rendered by the respective neural networks.

37. A method according to claim 36, wherein the networks diverse in architecture have differing window size and/or number of hidden units and/or number of output neurons.

38. A method according to claim 36, wherein the initial conditions are selected by the process of randomly setting each weight to .+-.0.1 and/or randomly selected from [-1; 1].

39. A method according to claim 36, wherein the learning set comprises sets generated from the X-fold cross-validation process.

40. A method according to claim 36, wherein the sub-types of input data are selected from the group comprising sequence profiles, amino acid composition, amino acid position and **peptide** length.

41. A method according to claim 36, wherein the sub-types of output data sets are selected from the group comprising secondary structure class assignment, tertiary structure, interatomic distance, bond strength, bond angle, descriptors relating to or reflecting hydrophobicity, hydrophilicity, acidity, basicity, relative nucleophilicity, relative electrophilicity, electron density or rotational freedom, scalar products of atomic vectors, cross products of atomic vectors, angles between atomic vectors, triple scalar products between atomic vectors, torsion angles, atomic angles such as but not exclusively omega, psi, phi, chi1, chi2, chi21, chi3, chi4, chi5 angles, chain curvature, chain torsion angles, and mathematical functions thereof.

42. A method according to claim 25, wherein the input data is taken unchanged or upon filtration through one or more quality filters from a

biological database, such as a **protein** database, a DNA data base and an RNA database.

43. A method according to claim 25, wherein the weighted networks outputs are averaged by a per-chain, per-subset of a chain, or per-residue confidence rating.

44. A method according to claim 43, wherein the per-residue confidence rating is calculated as the average per residue absolute difference between the highest probability and the second highest probability.

45. A method according to claim 43, wherein the per-subset of a chain confidence rating or per-chain confidence rating is calculated by multiplying each component of a single- or multi-component output for each residue, said output produced by the selected prediction means by the per-chain estimated accuracy obtained for said chain and prediction means, and the resulting products summed by residue and component, and the resulting sums being divided by the sum of weights, and the resulting maximal per-residue component quotient being used to determine the H or E or C secondary structure assignment for that residue, and the per-chain per-prediction probability in the H versus E versus C assignment is averaged over a given **protein** chain.

46. A method according to claim 25, wherein the output is a set number.

47. A method according to claim 25, wherein descriptors are selected from the group comprising secondary structure class assignment, tertiary structure, interatomic distance, bond strength, bond angle, descriptors relating to or reflecting hydrophobicity, hydrophilicity, acidity, basicity, relative nucleophilicity, relative electrophilicity, electron density or rotational freedom, scalar products of atomic vectors, cross products of atomic vectors, angles between atomic vectors, triple scalar products between atomic vectors, torsion angles, atomic angles such as but not exclusively omega, psi, phi, chi1, chi2, chi21, chi3, chi4, chi5 angles, chain curvature, chain torsion angles, torsion vectors and mathematical functions thereof.

48. A method according to claim 25, wherein a multi-component output comprises prediction with at least 2 components such as a 2-component, a 3-component, 4-component, or 5-component, or 10-component prediction.

49. A method according to claim 48, wherein a 3-component output comprises the prediction for a helix (H), an extended strand (E) and a coil (C).

50. A method according to claim 25, wherein the output of one level of neural networks comprises a descriptor of 2, 3, 4, 5, 6, 7, 8 or 9 consecutive residues, preferably 3, 5, 7, or 9 consecutive residues.

51. A method according to claim 25, wherein the number of neural networks in the one of the subsequent level or levels range from 1 to 1 000 000, such as from 1 to 100 000, 1 to 50 000, 1 to 10 000, 1 to 5000, 1 to 2500, 1 to 1000, 1 to 500, 1 to 250, 1 to 100, 1 to 50, 1 to 25 or 1 to 10.

52. A method of predicting a set of features of an input data by providing said input data to at least 16 diverse neural networks thereby providing an individual prediction of the said set of features on the basis of a weighted average said weighted average comprising an evaluation of the estimation of the prediction accuracy for a **protein** chain by a prediction means.

53. A method according to claim 52, wherein the estimation of the prediction accuracy is made by summing the per-residue maximum of H versus E versus C probabilities for said **protein** chain and dividing by the number of amino-acid residues in the **protein** chain, and wherein the mean and standard deviation of the accuracy estimation is taken for all prediction means for the **protein** chain, and wherein a weighted average is made for substantially all or optionally a subset of prediction means, wherein the subset comprises those prediction means with estimated accuracy above a threshold consisting of the mean estimated accuracy, the mean accuracy plus one standard deviation above the mean accuracy, or the mean estimated accuracy plus two standard deviations above the mean, or wherein the subset comprises at least 10 prediction means in cases where the accuracy of fewer than 10 estimated prediction fail to satisfy the threshold,

54. A method according to claim 52, wherein the weighted average comprise a multiplication of each component of a single- or multi-component output for each residue, said output produced by the selected prediction means by the per-chain estimated accuracy obtained for said chain and prediction means, and the resulting said products summed by residue and component, and the resulting sums being divided by the sum of weights, and the resulting maximal per-residue component quotient being used to determine the H or E or C secondary structure assignment for that residue, and the per-chain per-prediction probability in the H versus E versus C assignment is averaged over a given **protein** chain.

55. A method according to claim 52, wherein the set of features comprise secondary structure class assignment, tertiary structure, interatomic distance, bond strength, bond angle, descriptors relating to or reflecting hydrophobicity, hydrophilicity, acidity, basicity, relative nucleophilicity, relative electrophilicity, electron density or rotational freedom, scalar products of atomic vectors, cross products of atomic vectors, angles between atomic vectors, triple scalar products between atomic vectors, torsion angles, atomic angles such as but not exclusively omega, psi, phi, chi1, chi2, chi21, chi3, chi4, chi5 angles, chain curvature, chain torsion angles, torsion vectors and mathematical functions thereof.

56. A method according to claim 52, wherein the input data is provided to at least 20 diverse neural networks, such as at least 30, 40, 50, 60, 70, 80, 90, 100, 200, 500, 1000, 5000, 10 000, 100 000, and 1 000 000.

57. A method of predicting a set of features of input data using outputexpansion wherein a process by which a single- or multi-component output is represented by a descriptor of 2 or more consecutive elements of a sequence, such as residues of a **protein** sequence.

58. A method for predicting a set of chemical, physical or biological features related to chemical substances or related to interactions of chemical substances using a system comprising a prediction means comprising output expansion, the method comprising using at least 1 individual prediction means predicting substantially the whole set of features at least twice thereby providing at least two individual predictions of substantially all of the set of features, and predicting the set of features either on the basis of combining at least two of the individual predictions, the combining being performed in such a manner that the combined prediction is more accurate on a test set than substantially any of the at least two of the predictions, or on the basis of selecting one of the sets of predictions, the selection being performed in such a manner that the selected prediction is more accurate

on a test set than a prediction from corresponding prediction means without the use of output expansion, or predicting the set of features on the basis of at least one individual predictions, or on the basis of combining at least two of the individual predictions, the combining being performed in such a manner that the combined prediction is more accurate on a test set than substantially any of the predictions of the individual prediction means, or more accurate than corresponding prediction means not comprising output expansion.

L3 ANSWER 30 OF 73 USPATFULL
AN 2001:200372 USPATFULL
PI US 2001039480 A1 20011108
PI US 2001039480 A1 20011108
TI Apparatus and method for automated **protein** design
AB The present invention relates to apparatus and methods for quantitative protein design and optimization.
CLM What is claimed is:

1. A method executed by a computer under the control of a program, said computer including a memory for storing said program, said method comprising the steps of: (A) receiving a **protein** backbone structure with variable residue positions; (B) establishing a group of potential rotamers for each of said variable residue positions, wherein at least one variable residue position has rotamers from at least two different amino acid side chains; and (C) analyzing the interaction of each of said rotamers with all or part of the remainder of said **protein** backbone structure to generate a set of optimized **protein** sequences, wherein said analyzing step includes a Dead-End Elimination (DEE) computation.

2. A method executed by a computer under the control of a program, said computer including a memory for storing said program, said method comprising the steps of: (A) receiving a **protein** backbone structure with variable residue positions; (B) classifying each variable residue position as either a core, surface or boundary residue; (C) establishing a group of potential rotamers for each of said variable residue positions, wherein at least one variable residue position has rotamers from at least two different amino acid side chains; and (D) analyzing the interaction of each of said rotamers with all or part of the remainder of said **protein** to generate a set of optimized **protein** sequences.

3. A method according to claim 2 wherein said analyzing step comprises a DEE computation.

4. A method according to claim 1 or 2 wherein said set of optimized **protein** sequences comprises the globally optimal **protein** sequence.

5. A method according to claim 1 or 3 wherein said DEE computation is selected from the group consisting of original DEE and Goldstein DEE.

6. A method according to claim 1 or 2 wherein said analyzing step includes the use of at least one scoring function.

7. A method according to claim 6 wherein said scoring function is selected from the group consisting of a Van der Waals potential scoring function, a hydrogen bond potential scoring function, an atomic solvation scoring function, an electrostatic scoring function and a secondary structure propensity scoring function.

8. A method according to claim 6 wherein said analyzing step includes

the use of at least two scoring functions.

9. A method according to claim 6 wherein said analyzing step includes the use of at least three scoring functions.

10. A method according to claim 6 wherein said analyzing step includes the use of at least four scoring functions.

11. A method according to claim 6 wherein said atomic solvation scoring function includes a scaling factor that compensates for over-counting.

12. A method according to claim 1 or 2 further comprising testing at least one member of said set to produce experimental results.

13. A method according to claim 4 further comprising (D) generating a rank ordered list of additional optimal sequences from said globally optimal **protein** sequence.

14. A method according to claim 13 wherein said generating includes the use of a Monte Carlo search.

15. A method according to claim 2 wherein said analyzing step step comprises a Monte Carlo computation.

16. A method according to claim 13 further comprising: (E) testing some or all of said **protein** sequences from said ordered list to produce potential energy test results.

17. A method according to claim 16 further comprising: (F) analyzing the correspondence between said potential energy test results and theoretical potential energy data.

18. A method according to claim 1 or 2 further comprising altering at least one supersecondary structure parameter value of said **protein** backbone structure prior to establishing said potential rotamer group.

19. An optimized **protein** sequence generated by the method of claim 1 or 2.

20. A nucleic acid sequence encoding a **protein** sequence according to claim 19.

21. An expression vector comprising the nucleic acid of claim 20.

22. A host cell comprising the nucleic acid of claim 20.

23. A **protein** having a sequence that is at least about 5% different from a known **protein** sequence and is at least 20% more stable than the known **protein** sequence.

24. A computer readable memory to direct a computer to function in a specified manner, comprising: a side chain module to correlate a group of potential rotamers for residue positions of a **protein** backbone model; a ranking module to analyze the interaction of each of said rotamers with all or part of the remainder of said **protein** to generate a set of optimized **protein** sequences.

25. A computer readable memory according to claim 24 wherein said ranking module includes a van der Waals scoring function component.

26. A computer readable memory according to claim 24 wherein said

ranking module includes an atomic solvation scoring function component.

27. A computer readable memory according to claim 24 wherein said ranking module includes a hydrogen bond scoring function component.

28. A computer readable memory according to claim 24 wherein said ranking module includes a secondary structure scoring function component.

29. A computer readable memory according to claim 24 further comprising an assessment module to assess the correspondence between potential energy test results and theoretical potential energy data.

L3 ANSWER 42 OF 73 USPATFULL
AN 2001:122417 USPATFULL
PI US 6269312 B1 20010731
PI US 6269312 B1 20010731
TI Apparatus and method for automated **protein** design
AB The present invention relates to apparatus and methods for quantitative protein design and optimization.
CLM What is claimed is:

1. A method executed by a computer under the control of a program, said computer including a memory for storing said program, said method comprising the steps of: (A) receiving a **protein** backbone structure with variable residue positions; (B) altering at least one supersecondary structure parameter value of said **protein** backbone structure; (C) establishing a group of potential rotamers for each of said vahable residue positions, wherein the group of potential rotamers for at least one of said variable residue position has a rotamer selected from each of at least two diferent amino acid side chains; and (D) analyzing the interaction of each of said rotamers with all or part of the remainder of said **protein** backbone structure to generate a set of optimized **protein** sequences wherein said analyzing step includes a Dead-End Elimination (DEE) canmputation.

2. A method executed by a computer under the control of a program, said computer including a memory for storing said program, said method comprising the steps of: (A) receiving a **protein** backbone structure with variable residue positions; (B) altering at least one supersecondary structure parameter value of said **protein** backbone structure; (C) classifying each variable residue position as either a core, surface or boundary residue; (D) establishing a group of potential rotamers for each of said variable residue positions, wherein the group of potential rotaters for at least one of said variable residue position has a rotamer selected from each of at least two dilrerent amino acid side chains; and (E) analyzing the interaction of each of said rotamers with all or part of the remainder of said **protein** to generate a set of optimized **protein** sequences.

3. A method according to claim 2 wherein said analyzing step comprises a DEE computation.

4. A method according to claim 1 or 2 wherein said set of optimized **protein** sequences comprises the globally optimal **protein** sequence.

5. A method according to claim 1 or 3 wherein said DEE computation is selected from the group consisting of original DEE and Goldstein DEE.

6. A method according to claim 1 or 2 wherein said analyzing step includes the use of at least one scoring function.

7. A method according to claim 6 wherein said scoring function is selected from the group consisting of a Van der Waals potential scoring function, a hydrogen bond potential scoring function, an atomic solvation scoring function, an electrostatic scoring function and a secondary structure propensity scoring function.

8. A method according to claim 6 wherein said analyzing step includes the use of at least two scoring functions.

9. A method according to claim 6 wherein said analyzing step includes the use of at least three scoring functions.

10. A method according to claim 6 wherein said analyzing step includes the use of at least four scoring functions.

11. A method according to claim 6 wherein said atomic solvation scoring function includes a scaling factor that compensates for over-counting.

12. A method according to claim 1 or 2 further comprising experimentally testing at least one member of said set.

13. A method according to claim 4 further comprising the step of: generating a rank ordered list of additional optimal sequences from said globally optimal **protein** sequence.

14. A method according to claim 13 wherein said generating includes the use of a Monte Carlo search.

15. A method according to claim 2 wherein said analyzing step comprises a Monte Carlo computation.

16. A method according to claim 13 further comprising the step of: testing some or all of said **protein** sequences from said ordered list to produce potential energy test results.

17. A method according to claim 16 further comprising the step of: analyzing the correspondence between said potential energy test results and theoretical potential energy data.

18. A recombinant **protein** comprising an optimized **protein** sequence generated by the method of claim 1 or 2.

19. A nucleic acid sequence encoding a recombinant **protein** according to claim 18.

20. An expression vector comprising the nucleic acid sequence of claim 19.

21. A host cell comprising the nucleic acid sequence of claim 19.

22. A method executed by a computer under the control of a program, said computer including a memory for storing said program, said method comprising the steps of: (A) receiving a **protein** backbone structure with variable residue positions; (B) altering at least one supersecondary structure parameter value of said **protein** backbone structure; (C) establishing a group of potential rotamers for each of said variable residue positions, wherein the group of potential rotamers for at least one of said variable residue position has a rotamer selected from each of at least two diferent amino acid side

chains; and (D) analyzing the interaction of each of said rotamers with all or part of the remainder of said **protein** backbone structure to generate a set of optimized **protein** sequences, wherein said analyzing step includes: i. a Dead-End Elimination (DEE) computation; and, ii. at least one scoring function selected from the group consisting of a Van der Waals potential scoring function, a hydrogen bond potential scoring fanction, an atomic solvation scoring function, an electrostatic scoring function and a secondary structure propensity scoring function.

23. A method executed by a computer under the control of a program, said computer including a memory for storing said program, said method comprising the steps of: (A) receiving a **protein** backbone structure with variable residue positions; (B) altering at least one supersecondary structure parameter value of said **protein** backbone structure; (C) classifying each variable residue position as either a core, surface or boundary residue; (D) establishing a group of potential rotamers for each of said variable residue positions, wherein the group of potential rotamers for at least one of said variable residue position has a rotamer selected from each of at least two different amino acid side chains; and E) analyzing the interaction of each of said rotamers with all or part of the remainder of said **protein** to generate a set of optimized **protein** sequences wherein said analyzing step includes: i. a Dead-End Elimination (DEE) computation; and, ii. at least one scoring function selected from the group consisting of a Van der Waals potential scoring function, a hydrogen bond potential scoring function, an atomic solvation scoring function, an electrostatic scoring function and a secondary structure propensity scoring function.

L3 ANSWER 61 OF 73 USPATFULL
AN 1999:35118 USPATFULL
PI US 5884230 19990316
PI US 5884230 19990316
TI Method and system for **protein** modeling
AB A method in a computer system for modeling a three-dimensional structure of a model protein is provided. In one embodiment, the modeling is based upon a three-dimensional structure of a template protein and an amino acid sequence alignment of the model protein and the template protein. For each amino acid in the model protein, when the template protein has an amino acid aligned with the amino acid of the model protein, the position of the backbone atom of the amino acid of the model protein is established based on the position of a topologically equivalent backbone atom in the aligned amino acid of the template protein. In another embodiment, the modeling of a variable region of the model protein is based on a collection of .psi. and .phi. angle values for amino acid pairs in a family of proteins. In a further embodiment, these .psi. and .phi. angle values are classified according to a tetramer of adjacent amino acids and filtered based on a most probable conformation of portions of the variable region of the model protein.
CLM What is claimed is:
1. A method in a computer system for generating a collection of relative positional information between pairs of amino acids for use in modeling a three-dimensional structure of a variable region of a model **protein**, the computer system having relative positional information between pairs of amino acids in the variable regions of a collection of **proteins**, the method comprising: for each **protein** in the collection of **proteins**, for each pair of amino acids in the variable regions of the **protein**, in the collection of **proteins**, classifying the amino acids that are downstream and upstream from the amino acid pair; and storing the

relative positional information for the amino acid pair based on the amino acids in the amino acid pair and on the classification of the amino acids that are downstream and upstream so that the position of each amino acid in the variable region of the model **protein** can be modeled based on the stored relative positional information in the collection of **proteins** and based on classification of amino acids that are down stream and up stream from pairs of amino acids in the model **protein**.

2. The method of claim 1 wherein the relative positional information includes .psi. and .phi. angle values between pairs of amino acids.

3. The method of claim 1 wherein the amino acids are classified according to the amino acids that are adjacent to the amino acid pair.

4. The method of claim 3 wherein the adjacent amino acids are classified as charged, hydrophobic, and polar.

5. A method in a computer system for modeling a three-dimensional structure of a variable region of a model **protein**, the model **protein** having amino acids, the amino acids having positions within a three-dimensional structure, the method comprising the steps of: generating a collection of relative positional information between pairs of amino acids, the relative position information being classified by a tetramer of amino acids that include the amino acid pair; establishing a positional for a first amino acid of the variable region; and for each amino acid pair in the variable region, generating a model position for the amino acids based on a classification of the tetramer that includes the amino acid and based on the collection of relative positional information for the pairs of amino acids.

6. The method of claim 5 including the step of, for each combination of pairs of amino acids in variable regions of a family of **proteins** , collecting the .psi. and .phi. angle values for the pair of amino acids and wherein the step of generating a model position bases the model position on one of the collected .psi. and .phi. angle values.

7. The method of claim 6 wherein the step of generating a model position bases the model position on a randomly selected one of the collected .psi. and .phi. angle values.

8. The method of claim 5 including the step of generating a model position for the amino acids of the adjacent structurally conserved regions based on relative position information between pairs of amino acids.

9. The method of claim 8 wherein the step of generating a model position for the amino acids of the adjacent structurally conserved region is based on the .psi. and .phi. angle values in a corresponding region of a template **protein**.

10. The method of claim 8 including the step of comparing the generated model position for the amino acids of the adjacent structurally conserved region to positions in a corresponding region in a template **protein** to indicate effectiveness of the modeling.

11. A method in a computer system for modeling a three-dimensional structure of a variable region of a model **protein**, the model **protein** having amino acids, the method comprising the step of establishing positional information for the amino acids in the variable region based on .psi. and .phi. angle values between pairs of amino acids in a collection of **proteins** such that the pairs of amino

acids are classified according to their downstream and upstream amino acids.

12. The method of claim 11 wherein the model **protein** has a structurally conserved region that is adjacent to the variable region, and including the steps of: establishing positional information for the amino acids in the adjacent structurally conserved region based on the established positional information of the amino acids in the variable region and based on .psi. and .phi. angle values in a corresponding structurally conserved region of the template **protein**; and comparing the established positional information of the amino acids of the adjacent structurally conserved region to positional information of the corresponding structurally conserved region of the template **protein** to measure the effectiveness of the modeling.

13. The method of claim 11 including the step of collecting .psi. and .phi. angle values for pairs of amino acids in a template **protein** and classifying the collected .psi. and .phi. values based on the adjacent downstream and upstream amino acids and wherein the step of establishing bases the positional information on the collected .psi. and .phi. angle values.

14. The method of claim 13 wherein the step of establishing includes the step of randomly selecting a collected .psi. and .phi. angle value.

15. The method of claim 13 wherein the step of collecting collects .psi. and .phi. angle values for pairs of amino acids in a family of template **proteins**.

16. A method in a computer system for modeling a three-dimensional structure for a variable region of a model **protein**, the **protein** having amino acids, the variable region having a corresponding beginning structurally conserved region and a corresponding ending structurally conserved region, the method comprising the steps of: collecting .psi. and .phi. angle values for pairs of amino acids in a family of template **proteins** and classifying each .psi. and .phi. angle value according to the amino acids in the pair and an amino acid that is adjacent to the pair; generating three-dimensional positional information for the amino acids in the beginning structurally conserved region; generating three-dimensional positional information for the amino acids in the variable region based on the collected .psi. and .phi. angle values and based on the generated positional information for the beginning structurally conserved region; generating three-dimensional positional information for the amino acids in the ending structurally conserved region based on the generated positional information for the variable region and based on positional information for a corresponding ending structurally conserved region in a template **protein**; and comparing the generated positional information for the amino acids in the ending structurally conserved regions to positional information for the amino acids in the corresponding structurally conserved region in the template **protein** to indicate correctness of the model.

17. The method of claim 16 including the step of randomly selecting collected .psi. and .phi. angle values when generating the positional information for the amino acids in the variable region.

18. The method of claim 16 including the step of repeating the steps of generating and comparing, and including the step of selecting, as the model of the variable region, generated positional information for the amino acids in the variable region when the generated positional information of the amino acids in the ending structurally conserved

region most closely compares to positional information in the corresponding structurally conserved region in the template **protein**.

L3 ANSWER 62 OF 73 USPATFULL
AN 1999:28814 USPATFULL
PI US 5878373 19990302
PI US 5878373 19990302
TI System and method for determining three-dimensional structure of **protein** sequences
AB The present invention pertains to a system and method for predicting the protein fold of a target amino acid residue sequence of unknown protein structure. A target sequence is represented by a sequence of residue variability types that utilizes positional variability information present in an associated family of homologous sequences to the target sequence. The use of the positional variability information increases the likelihood of matching the target sequence with a known protein structure. In a first preferred embodiment, a target sequence is mapped into a sequence of residue variability types that are based on the solubility variability present between amino acid residues in homologous sequences. In a second preferred embodiment, each residue variability type represents a cluster of residue types at each position of aligned sets of homologous protein sequences. Each distinct cluster represents a pattern of residue variability at various positions in sets of homologous protein sequences. The sequence of residue variability types is aligned with one or more environment strings, each of which represents a known protein structure in accordance with the degree of surface exposure for each amino acid position in the protein's structure. The alignment is performed using a threading procedure that determines a score for each alignment indicating the compatibility of the sequence to the structure. The protein structure associated with the highest score is deemed to be the most analogous structure to the target sequence.
CLM What is claimed is:
1. A computer-implemented method of characterizing a **protein** sequence's three-dimensional structure, comprising the steps of: establishing access by a digital data processor to a database of **protein** sequences having known three-dimensional structures, each **protein** sequence comprising a sequence of residues, said database including for each **protein** sequence of known structure a corresponding sequence of residue environment values, each residue environment value representing at least one structural characteristic associated with at least one residue in said sequence of residues; using the digital data processor: for a given input **protein** sequence of unknown three-dimensional structure, identifying a set of homologous **protein** sequences; generating for said given input **protein** sequence and said homologous **protein** sequences, a corresponding sequence of residue variability types, wherein each residue variability type is selected from a defined set of variability types, each representing a respective positional variability measure of residues associated with various sequence positions in the input and homologous **protein** sequences; for each of at least a subset of the **protein** sequences in said database, selecting an alignment of said generated sequence of residue variability types yielding a highest score in accordance with a predefined scoring method and associating with each said **protein** sequence in said subset a match score ~~corresponding to said highest score; selecting~~ a **protein** structure associated with a **protein** sequence in said database having a highest match score; and outputting to at least one output device ~~information~~ identifying the selected **protein** structure.

2. The method of claim 1, said predefined scoring method assigns scores to every defined residue variability type with respect to every defined residue environment value, each score indicating a relative probability that a residue of a respective residue variability type will be found in a portion of any **protein** structure assigned a respective residue environment value.

3. The method of claim 1, said structural characteristic indicating a degree of exterior surface area exposure of an associated residue in a corresponding **protein** sequence.

4. The method of claim 3, said residue environment values selected from the set consisting of exposed, buried, and partially buried.

5. The method of claim 1, said positional variability measure includes a solubility variability measure.

6. The method of claim 1, said generating step further comprising the steps of: analyzing solubility variations between residues in similar sequence positions in said given input **protein** sequence and said selected homologous **protein** sequences; and associating each said solubility variation with a corresponding residue variability type.

7. The method of claim 6, said solubility variations including hydrophobic variability, hydrophobic invariability, hydrophilic variability, and hydrophilic invariability.

8. The method of claim 6, said analyzing step further comprising the steps of: determining a hydrophobic variability factor and a hydrophilic variability factor for each said residue position, said hydrophobic variability factor determined in accordance with the following mathematical relation: ##EQU5## said hydrophilic variability factor determined in accordance with the following mathematical relation: ##EQU6## where N is the number of sequences, i is a residue position, n.sub.ik is the ith amino acid of the kth sequence, d.sub.kl is a measure of evolutionary distance between the kth and lth sequences, w.sub.k and w.sub.l are weights associated with the kth and lth sequence, H.phi. is a set of hydrophobic amino acids residues including {Phe, Ile, Leu, Met, Val, Trp}, HP is a set of hydrophilic amino acids residues including {Asp, Glu, Lys, Asn, Gln, Arg, Ser}, and HA is a set of ambivalent amino acids residues including {Ala, Cys, Gly, His, Pro, Thr, and Tyr}; said associating step further comprising the steps of: classifying each said residue position in accordance with one of the following classifications: hydrophobic variant, if said hydrophobic variability factor>A, hydrophobic invariant, if said hydrophobic variability factor<A; and classifying each said residue position in accordance with one of the following classifications: hydrophilic variant, if said hydrophilic variability factor>B, hydrophilic invariant, if said hydrophilic variability factor<B; wherein A and B are median hydrophobic and hydrophilic variability factors.

9. The method of claim 6, said solubility variations including hydrophilic variability, hydrophilic invariability, and hydrophilic partially variant.

10. The method of claim 1, said residue variability types including each amino acid residue classified in accordance with each of four classes selected from the set consisting of (hydrophobic variant, hydrophilic variant), (hydrophobic variant, hydrophilic invariant), (hydrophobic invariant, hydrophilic variant), and (hydrophobic invariant, hydrophilic

invariant).

11. The method of claim 1, said residue variability types including each amino acid residue classified in accordance with each of three classes selected from the set consisting of hydrophilic variant, hydrophilic invariant, and hydrophilic partially variant.

12. The method of claim 1, said positional variability measure based on a cluster analysis of residue positional variability within a set of multiple sequence alignments corresponding to known **protein** structures.

13. The method of claim 1, said generating step further comprising the steps of: providing a plurality of cluster vectors, each said cluster vector associated with a particular residue variability type; determining a residue vector for each said residue position in said given input **protein** sequence and said selected homologous **protein** sequences, each said residue vector indicating a frequency of occurrence of each residue within said residue position; matching each said residue vector with a closest cluster vector; and representing each said residue position with a residue variability type associated with said matched cluster vector.

14. A computer-implemented method for characterizing a **protein** sequence's three-dimensional structure, comprising the steps of: providing a target sequence of residues of unknown three-dimensional structure; using a digital data processor: identifying a set of homologous **protein** sequences for said target sequence; mapping said target sequence and said set of homologous **protein** sequences to a corresponding first sequence of residue variability types, each said residue variability type of said first sequence associated with a first positional variability measure; mapping said target sequence and said set of homologous **protein** sequences to a corresponding second sequence of residue variability types, each said residue variability type of said second sequence associated with a second positional variability measure; determining a first predicted **protein** structure for said first sequence of residue variability types and a second predicted **protein** structure for said second sequence of residue variability types; utilizing said predicted **protein** structures to determine an analogous **protein** structure to said target sequence; and outputting to at least one output device information identifying the analogous **protein** structure.

15. The method of claim 14, providing a plurality of environment strings, each of said environment strings characterizing a **protein** structure as a sequence of environment classes, each said environment class representing at least one structural characteristic associated with at least one residue in said corresponding **protein** structure; and said determining step further comprising the steps of: comparing each of said environment strings with said first sequence of residue variability types in order to determine a first predicted **protein** structure; and comparing each of said environment strings with said second sequence of residue variability types in order to determine a second predicted **protein** structure.

16. The method of claim 15, said structural characteristic corresponding to degree of exterior surface area exposure of an associated residue in a corresponding **protein** structure.

17. The method of claim 16, said environment classes selected from the

set consisting of exposed, buried, and partially buried.

18. The method of claim 14, said utilizing step further comprising the step of: when said first and second predicted **protein** structures differ, performing a structural comparison of said predicted **protein** structures, said structural comparison generating a similarity measure indicating structural similarity of said first and second predicted **protein** structures.

19. The method of claim 18, reporting each said predicted **protein** structure and a respective confidence level, wherein said reported confidence level for each reported predicted **protein** structure is substantially higher when said first and second predicted **protein** structures match than when said first and second predicted **protein** structures differ.

20. The method of claim 14, said first positional variability measure including a solubility variability measure of residues associated with various sequence positions in said target sequence and said set of homologous **protein** sequences.

21. The method of claim 14, said second positional variability measure including a respective measure based on a cluster analysis of residue positional variability within a set of multiple sequence alignments of known **protein** structures.

22. A computer system for characterizing a **protein** sequence's three-dimensional structure, said system comprising of: a memory for storing a database of **protein** structures, each of said **protein** structures having a corresponding sequence of residue environment values, each residue environment value representing at least one structural characteristic associated with at least one residue in one of said **protein** structures, each of said **protein** structures having a corresponding **protein** sequence of residues, a set of residue variability types, each of said residue variability types representing a respective positional variability measure of residues associated with various sequences positions in said **protein** sequences, a given input **protein** sequence of unknown three-dimensional structure; and a **protein** structure determination procedure including instructions for identifying a set of homologous **protein** sequences for said given input **protein** sequence, converting said given input sequence and said homologous sequences into a corresponding sequence of residue variability types, selecting a best alignment of said sequence of residue variability types with each of at least a subset of said **protein** structures in said database, including generating a respective match score for said best alignment of said sequence of residue variability types with each of said subset of said **protein** structures in said database, and select a **protein** structure associated with a **protein** sequence in said database having a highest match score.

23. The system of claim 22, said structural characteristic indicating a degree of exterior surface area exposure of an associated residue in a corresponding **protein** sequence.

24. The system of claim 23, said residue environment values selected from the set consisting of exposed, buried, and partially buried.

25. The system of claim 24, said positional variability measure includes a solubility variability measure.

26. The system of claim 22, said instructions for converting in said **protein** determination procedure including instructions to analyze solubility variations between residues in similar sequence positions in said given input **protein** sequence and said selected homologous **protein** sequences, and associate said solubility variations for each input **protein** sequence position with a corresponding residue variability type.

27. The system of claim 26, said solubility variations including hydrophobic variability, hydrophobic invariability, hydrophilic variability, and hydrophilic invariability.

28. The system of claim 26, said **protein** structure determination procedure further including instructions to determine a hydrophobic variability factor and a hydrophilic variability factor for each said sequence position in said given input **protein** sequence and said selected homologous sequences, said hydrophobic variability factor determined in accordance with the following mathematical relation: ##EQU7## said hydrophilic variability factor determined in accordance with the following mathematical relation: ##EQU8## where N is the number of sequences, i is a residue position, n.sub.ik is the ith amino acid of the kth sequence, d.sub.kl is a measure of the evolutionary distance between the kth and lth sequences, w.sub.k and w.sub.l are weights associated with the kth and lth sequence, H.phi. is a set of hydrophobic amino acids residues including {Phe, Ile, Leu, Met, Val, Trp}, HP is a set of hydrophilic amino acids residues including {Asp, Glu, Lys, Asn, Gln, Arg, Ser}, and HA is a set of ambivalent amino acids residues including {Ala, Cys, Gly, His, Pro, Thr, and Tyr}; classify each said residue position of said given input **protein** sequence in accordance with one of the following classifications: hydrophobic variant, if said hydrophobic variability factor>A, hydrophobic invariant, if said hydrophobic variability factor<A; and classify each said residue position of said given input **protein** sequence in accordance with one of the following classifications: hydrophilic variant, if said hydrophilic variability factor>B, hydrophilic invariant, if said hydrophilic variability factor<B; wherein A and B are median hydrophobic and hydrophilic variability factors.

29. The system of claim 22, said residue variability types including each amino acid residue classified in accordance with each of four classes selected from the set consisting of (hydrophobic variant, hydrophilic variant), (hydrophobic variant, hydrophilic invariant), (hydrophobic invariant, hydrophilic variant), and (hydrophobic invariant, hydrophilic invariant).

30. The system of claim 22, said residue variability types including each amino acid residue classified in accordance with each of three classes selected from the set consisting of hydrophilic variant, hydrophilic invariant, and hydrophilic partially variant.

31. The system of claim 22, said positional variability measure is based on a cluster analysis of residue positional variability within a set of multiple sequence alignments for known **protein** structures.

32. The system of claim 22, said converting instructions in said **protein** structure determination procedure including instructions for providing a plurality of cluster vectors, each said cluster vector associated with a particular residue variability type and representing a pattern of residue variability at various positions in sets of homologous **protein** sequences, determining a residue vector for each said residue position in said given input **protein**

sequence and said selected homologous **protein** sequences, said residue vector indicating a frequency of occurrence of distinct residues in said residue position, matching each said residue vector with a closest cluster vector, and representing each said residue position with one of said residue variability types associated with said matched cluster vector.

33. A computer system for characterizing a **protein** sequence's three-dimensional structure, comprising: a memory for storing a target sequence of residues of unknown three-dimensional structure, a set of homologous **protein** sequences for said target sequence, a first set of residue variability types, each said residue variability type of said first set associated with a first positional variability measure, a second set of residue variability types, each said residue variability type of said second set associated with a second positional variability measure; a **protein** structure determination procedure including instructions that select one of said sets of residue variability types, map said target sequence and said set of homologous **protein** sequences to a sequence of said selected residue variability types, and determine a predicted **protein** structure for said sequence of residue variability types; and a structural comparison procedure for comparing any two specified **protein** structures, said structural comparison procedure including instructions to generate a similarity measure indicative of structural similarity of said two specified **protein** structures; wherein said system executes said **protein** structure determination procedure a first time utilizing said first set of residue variability types and generating a first predicted **protein** structure, executes said **protein** structure determination procedure a second time utilizing said second set of residue variability types and generating a second predicted **protein** structure, and executes said structural comparison procedure when said first predicted **protein** structure and said second predicted **protein** structure differ to generate a measure of structural similarity of said first and second predicted **protein** structures.

34. The system of claim 33, including a reporting procedure that reports each predicted **protein** structure and a respective confidence level.

35. The system of claim 33, said structural comparison procedure including instructions to quantify topological differences between said predicted **protein** structures.

36. The system of claim 33, said first positional variability measure including a solubility variability measure of residues associated with various sequence positions in said target sequence and said set of homologous **protein** sequences.

37. The system of claim 33, said second positional variability measure including a respective measure based on a cluster analysis of residue positional variability within a set of multiple sequence alignments of known **protein** structures.

38. A computer program product for use in conjunction with a computer system, the computer program product comprising a computer readable storage medium and a computer program mechanism embedded therein, the computer program mechanism comprising: a database of **protein** structures, each of said **protein** structures having a corresponding sequence of residue environment values, each residue environment value representing at least one structural characteristic associated with at least one residue in one of said **protein**

structures, each of said **protein** structures having a corresponding **protein** sequence of residues; instructions for storing a given input **protein** sequence of unknown three-dimensional structure; and a **protein** structure determination procedure including instructions for identifying a set of homologous **protein** sequences for said given input **protein** sequence, converting said given input sequence and said homologous sequences into a corresponding sequence of residue variability types, each of said residue variability types selected from a predefined set of residue variability types, each of said residue variability types representing a respective positional variability measure of residues associated with various sequences positions in said **protein** sequences, selecting a best alignment of said sequence of residue variability types with each of at least a subset of said **protein** structures in said database, including generating a respective match score for said best alignment of said sequence of residue variability types with each of said subset of said **protein** structures in said database, and select a **protein** structure associated with a **protein** sequence in said database having a highest match score.

39. The computer program product of claim 38, said structural characteristic indicating a degree of exterior surface area exposure of an associated residue in a corresponding **protein** sequence.

40. The computer program product of claim 38, said residue environment values selected from the set consisting of exposed, buried, and partially buried.

41. The computer program product of claim 38, said positional variability measure includes a solubility variability measure.

42. The computer program product of claim 38, said instructions for converting in said **protein** determination procedure including instructions to analyze solubility variations between residues in similar sequence positions in said given input **protein** sequence and said selected homologous **protein** sequences, and associate said solubility variations for each input **protein** sequence position with a corresponding residue variability type.

43. The computer program product of claim 38, said solubility variations including hydrophobic variability, hydrophobic invariability, hydrophilic variability, and hydrophilic invariability.

44. The computer program product of claim 38, said **protein** structure determination procedure further including instructions to determine a hydrophobic variability factor and a hydrophilic variability factor for each said sequence position in said given input **protein** sequence and said selected homologous sequences, said hydrophobic variability factor determined in accordance with the following mathematical relation: ##EQU9## said hydrophilic variability factor determined in accordance with the following mathematical relation: ##EQU10## where N is the number of sequences, i is a residue position, n.sub.ik is the ith amino acid of the kth sequence, d.sub.kl is a measure of the evolutionary distance between the kth and lth sequences, w.sub.k and w.sub.l are weights associated with the kth and lth sequence, H.phi. is a set of hydrophobic amino acids residues including {Phe, Ile, Leu, Met, Val, Trp}, HP is a set of hydrophilic amino acids residues including {Asp, Glu, Lys, Asn, Gln, Arg, Ser}, and HA is a set of ambivalent amino acids residues including {Ala, Cys, Gly, His, Pro, Thr, and Tyr}; classify each said residue position of said given input **protein** sequence in accordance with one of the

following classifications: hydrophobic variant, if said hydrophobic variability factor>A, hydrophobic invariant, if said hydrophobic variability factor<A; and classify each said residue position of said given input **protein** sequence in accordance with one of the following classifications: hydrophilic variant, if said hydrophilic variability factor>B, hydrophilic invariant, if said hydrophilic variability factor<B; wherein A and B are median hydrophobic and hydrophilic variability factors.

45. The computer program product of claim 38, said residue variability types including each amino acid residue classified in accordance with each of four classes selected from the set consisting of (hydrophobic variant, hydrophilic variant), (hydrophobic variant, hydrophilic invariant), (hydrophobic invariant, hydrophilic variant), and (hydrophobic invariant, hydrophilic invariant).

46. The computer program product of claim 38, said residue variability types including each amino acid residue classified in accordance with each of three classes selected from the set consisting of hydrophilic variant, hydrophilic invariant, and hydrophilic partially variant.

47. The computer program product of claim 38, said positional variability measure is based on a cluster analysis of residue positional variability within a set of multiple sequence alignments for known **protein** structures.

48. The computer program product of claim 38, said converting instructions in said **protein** structure determination procedure including instructions for providing a plurality of cluster vectors, each said cluster vector associated with a particular residue variability type and representing a pattern of residue variability at various positions in sets of homologous **protein** sequences, determining a residue vector for each said residue position in said given input **protein** sequence and said selected homologous **protein** sequences, said residue vector indicating a frequency of occurrence of distinct residues in said residue position, matching each said residue vector with a closest cluster vector, and representing each said residue position with one of said residue variability types associated with said matched cluster vector.

49. A computer program product for use in conjunction with a computer system, the computer program product comprising a computer readable storage medium and a computer program mechanism embedded therein, the computer program mechanism comprising: instructions for storing a target sequence of residues of unknown three-dimensional structure; instructions for identifying a set of homologous **protein** sequences for said target sequence; a **protein** structure determination procedure including instructions that selects one of at least two sets of residue variability types, said at least two sets of residue variability types including a first set of residue variability types, each said residue variability type of said first set associated with a first positional variability measure, and a second set of residue variability types, each said residue variability type of said second set associated with a second positional variability measure; maps said target sequence and said set of homologous **protein** sequences to a sequence of residue variability types from said selected set of residue variability types, and determines a predicted **protein** structure for said sequence of residue variability types; and a structural comparison procedure for comparing any two specified **protein** structures, said structural comparison procedure including instructions to generate a similarity measure indicative of structural similarity of said two specified **protein**

structures; wherein said system executes said **protein** structure determination procedure a first time utilizing said first set of residue variability types and generating a first predicted **protein** structure, executes said **protein** structure determination procedure a second time utilizing said second set of residue variability types and generating a second predicted **protein** structure, and executes said structural comparison procedure when said first predicted **protein** structure and said second predicted **protein** structure differ to generate a measure of structural similarity of said first and second predicted **protein** structures.

50. The computer program product of claim 49, including a reporting procedure that reports each predicted **protein** structure and a respective confidence level.

51. The computer program product of claim 49, said structural comparison procedure including instructions to quantify topological differences between said predicted **protein** structures.

52. The computer program product of claim 49, said first positional variability measure including a solubility variability measure of residues associated with various sequence positions in said target sequence and said set of homologous **protein** sequences.

53. The computer program product of claim 49, said second positional variability measure including a respective measure based on a cluster analysis of residue positional variability within a set of multiple sequence alignments of known **protein** structures.

L3 ANSWER 63 OF 73 USPATFULL
AN 1998:148842 USPATFULL
PI US 5842151 19981124
PI US 5842151 19981124
TI Prediction method and apparatus for a secondary structure of **protein**
AB A prediction method and apparatus for secondary structures of a protein in which accuracy in prediction of a formation of a .beta.-sheet is increased and which can applied to any type of protein including an .alpha.-helix and a .beta.-sheet. Formation of the .alpha.-helix is predicted with respect to each amino acid residue in a sequence of amino acid residues. Then, formation of the .beta.-sheet is predicted with respect to all pairs of residues which were not predicted to form the .alpha.-helix. Results of prediction of the .alpha.-helix and the .beta.-sheet are combined to obtain a result of prediction of the secondary structure of the protein.
CLM What is claimed is:
1. A method for predicting secondary structures which are characteristic structures of a **protein** including an .alpha.-helix and a .beta.-sheet, comprising the steps of: a) predicting a formation of the .alpha.-helix with respect to each amino acid residue in a sequence; b) predicting a formation of the .beta.-sheet with respect to all pairs of amino acid residues which are not predicted to form the .alpha.-helix by step a); and c) combining results obtained by step a) and step b) to obtain a result of prediction of the secondary structures of the **protein**, wherein the step a) comprises the steps of: a-1) learning which types of amino acid residues have a tendency to form the .alpha.-helix; a-2) determining the formation of the .alpha.-helix with respect to each amino acid residue in said sequence based on results obtained by step a-1); and a-3) providing a mark to each amino acid residue which was determined to form the .alpha.-helix by step a-2), and

subjecting all of the amino acid residues determined not to form the .alpha.-helix by step a-2) to the prediction of the .beta.-sheet, and wherein a determination of step a-2) is made based on consecutiveness of the amino acid residues having a predetermined level of formability of the .alpha.-helix.

2. The method as claimed in claim 1, wherein the consecutiveness of the amino acid residues is determined based on a series of amino acid residues comprising four amino acid residues.

3. A method for predicting secondary structures which are characteristic structures of a **protein** including an .alpha.-helix and a .beta.-sheet, comprising the steps of: a) predicting a formation of the .alpha.-helix with respect to each amino acid residue in a sequence; b) predicting a formation of the .beta.-sheet with respect to all pairs of amino acid residues which are not predicted to form the .alpha.-helix by step a); and c) combining results obtained by step a) and step b) to obtain a result of prediction of the secondary structures of the **protein**, wherein the step b) comprises the steps of: b-1) determining a .beta.-sheet tendency index with respect to all pairs of amino acid residues which were not predicted to form the .alpha.-helix by the step a); b-2) selecting candidate amino acid residues forming the .beta.-sheet by comparing the .beta.-sheet tendency index with a predetermined threshold value, amino acid residues of a pair having a .beta.-sheet tendency index greater than the threshold value being selected as the candidate amino acid residues; and b-3) seeking a series of candidate amino acid residues comprising a maximum number of candidate amino acid residues from among the candidate residues selected by step b-2) so that said series of candidate amino acid residues is determined to form the .beta.-sheet.

4. The method as claimed in claim 3, wherein in step b-3), when less than a predetermined number of consecutive amino acid residues is not selected as the candidate amino acid residues, said non-selected consecutive amino acid residues are regarded as the amino acid residues forming the .beta.-sheet.

5. An apparatus for predicting secondary structures which are characteristic structures of a **protein** including an .alpha.-helix and a .beta.-sheet, said apparatus comprising: .alpha.-helix predicting means for predicting a formation of the .alpha.-helix with respect to each amino acid residue in a sequence; .beta.-sheet predicting means for predicting a formation of the .beta.-sheet with respect to all pairs of amino acid residues which were not predicted to form the .alpha.-helix by said .alpha.-helix predicting means; and combining means for combining results obtained by said .alpha.-helix predicting means and said .beta.-sheet predicting means to obtain a result of prediction of the secondary structure of the **protein**, wherein said .alpha.-helix predicting means comprises: learning means for learning which types of amino acid residues have a tendency to form the .alpha.-helix; determining means for determining the formation of the .alpha.-helix with respect to each amino acid residue in said sequence based on results obtained by said learning means; and providing means for providing a mark to each residue which was determined to form the .alpha.-helix by said determining means, and subjecting all of the residues determined not to form the .alpha.-helix by said determining means to the prediction of the .beta.-sheet; and wherein a determination by said determining means is made based on consecutiveness of the residues having a predetermined level of formability of the .alpha.-helix.

6. An apparatus for predicting secondary structures which are

characteristic structures of a **protein** including an .alpha.-helix and a .beta.-sheet, said apparatus comprising: .alpha.-helix predicting means for predicting a formation of the .alpha.-helix with respect to each amino acid residue in a sequence; .beta.-sheet predicting means for predicting a formation of the .beta.-sheet with respect to all pairs of amino acid residues which were not predicted to form the .alpha.-helix by said .alpha.-helix predicting means; and combining means for combining results obtained by said .alpha.-helix predicting means and said .beta.-sheet predicting means to obtain a result of prediction of the secondary structure of the **protein**, wherein said .alpha.-helix predicting means comprises: learning means for learning which types of amino acid residues have a tendency to form the .alpha.-helix; determining means for determining the formation of the .alpha.-helix with respect to each amino acid residue in said sequence based on results obtained by said learning means; and providing means for providing a mark to each residue which was determined to form the .alpha.-helix by said determining means, and subjecting all of the residues determined not to form the .alpha.-helix by said determining means to the prediction of the .beta.-sheet; wherein a determination by said determining means is made based on consecutiveness of the residues having a predetermined level of formability of the .alpha.-helix, and wherein the consecutiveness of the residues is determined based on a series of residues comprising four residues.

7. An apparatus for predicting secondary structures which are characteristic structures of a **protein** including an .alpha.-helix and a .beta.-sheet, said apparatus comprising: .alpha.-helix predicting means for predicting a formation of the .alpha.-helix with respect to each amino acid residue in a sequence; .beta.-sheet predicting means for predicting a formation of the .beta.-sheet with respect to all pairs of amino acid residues which were not predicted to form the .alpha.-helix by said .alpha.-helix predicting means; and combining means for combining results obtained by said .alpha.-helix predicting means and said .beta.-sheet predicting means to obtain a result of prediction of the secondary structure of the **protein**, wherein said .beta.-sheet predicting means comprises: determining means for determining a .beta.-sheet tendency index with respect to all pairs of residues which were not predicted to form the .alpha.-helix by said .alpha.-helix predicting means; selecting means for selecting candidate residues forming the .beta.-sheet by comparing the .beta.-sheet tendency index with a predetermined threshold value, residues of a pair having a .beta.-sheet tendency index greater than the threshold value being selected as the candidate residues; and seeking means for seeking a series of candidate residues comprising a maximum number of candidate residues from among the candidate residues selected by said selecting means so that said series of candidate residues is determined to form the .beta.-sheet.

8. The apparatus as claimed in claim 7, wherein when less than a predetermined number of consecutive residues is not selected as the candidate residues, said non-selected consecutive residues are regarded as the residues forming the .beta.-sheet.

9. A method for predicting secondary structures which are characteristic structures of a **protein** including an .alpha.-helix and a .beta.-sheet, comprising the steps of: a) predicting a formation of the .alpha.-helix with respect to each amino acid residue in a sequence; b) predicting a formation of the .beta.-sheet with respect to all pairs of amino acid residues which are not predicted to form the .alpha.-helix by step a); and c) combining results obtained by step a) and step b) to obtain a result of prediction of the secondary structures of the

protein, wherein the step a) includes the steps of: a-1) learning which types of amino acid residues have a tendency to form the .alpha.-helix; a-2) determining the formation of the .alpha.-helix with respect to each amino acid residue in said sequence based on results obtained by step a-1); and a-3) providing a mark to each amino acid residue which was determined to form the .alpha.-helix by step a-2), and subjecting all of the amino acid residues determined not to form the .alpha.-helix by step a-2) to the prediction of the .beta.-sheet, and wherein the step b) includes the steps of: b-1) determining a .beta.-sheet tendency index with respect to all pairs of amino acid residues which were not predicted to form the .alpha.-helix by the step a); b-2) selecting candidate amino acid residues forming the .beta.-sheet by comparing the .beta.-sheet tendency index with a predetermined threshold value, amino acid residues of a pair having a .beta.-sheet tendency index greater than the threshold value being selected as the candidate amino acid residues; and b-3) seeking a series of candidate amino acid residues comprising a maximum number of candidate amino acid residues from among the candidate residues selected by step b-2) so that said series of candidate amino acid residues is determined to form the .beta.-sheet.

10. The method as claimed in claim 9, wherein a determination of step a-2) is made based on consecutiveness of the amino acid residues having a predetermined level of formability of the .alpha.-helix.

11. The method as claimed in claim 10, wherein the consecutiveness of the amino acid residues is determined based on a series of amino acid residues comprising four amino acid residues.

12. The method as claimed in claim 9, wherein in step b-3), when less than a predetermined number of consecutive amino acid residues is not selected as the candidate amino acid residues, said non-selected consecutive amino acid residues are regarded as the amino acid residues forming the .beta.-sheet.

13. An apparatus for predicting secondary structures which are characteristic structures of a **protein** including an .alpha.-helix and a .beta.-sheet, said apparatus comprising: .alpha.-helix predicting means for predicting a formation of the .alpha.-helix with respect to each amino acid residue in a sequence; .beta.-sheet predicting means for predicting a formation of the .beta.-sheet with respect to all pairs of amino acid residues which were not predicted to form the .alpha.-helix by said .alpha.-helix predicting means; and combining means for combining results obtained by said .alpha.-helix predicting means and said .beta.-sheet predicting means to obtain a result of prediction of the secondary structure of the **protein**, wherein said .alpha.-helix predicting means comprises: learning means for learning which types of amino acid residues have a tendency to form the .alpha.-helix; determining means for determining the formation of the .alpha.-helix with respect to each amino acid residue in said sequence based on results obtained by said learning means; and providing means for providing a mark to each residue which was determined to form the .alpha.-helix by said determining means, and subjecting all of the residues determined not to form the .alpha.-helix by said determining means to the prediction of the .beta.-sheet, and wherein said .beta.-sheet predicting means comprises: determining means for determining a .beta.-sheet tendency index with respect to all pairs of residues which were not predicted to form the .alpha.-helix by said .alpha.-helix predicting means; selecting means for selecting candidate residues forming the .beta.-sheet by comparing the .beta.-sheet tendency index with a predetermined threshold value, residues of a pair having a .beta.-sheet tendency index greater than the threshold value being

selected as the candidate residues; and seeking means for seeking a series of candidate residues comprising a maximum number of candidate residues from among the candidate residues selected by said selecting means so that said series of candidate residues is determined to form the .beta.-sheet.

14. The apparatus as claimed in claim 13, wherein a determination by said determining means is made based on consecutiveness of the residues having a predetermined level of formability of the .alpha.-helix.

15. The apparatus as claimed in claim 14, wherein the consecutiveness of the residues is determined based on a series of residues comprising four residues.

16. The apparatus as claimed in claim 13, wherein when less than a predetermined number of consecutive residues is not selected as the candidate residues, said non-selected consecutive residues are regarded as the residues forming the .beta.-sheet.

L3 ANSWER 65 OF 73 USPATFULL
AN 97:23331 USPATFULL
PI US 5612895 19970318
PI US 5612895 19970318
TI Method of rational drug design based on ab initio computer simulation of conformational features of **peptides**
AB A method of rational drug design includes simulating polypeptides in a way that predicts the most probable secondary and/or tertiary structures of a polypeptide, e.g., an oligopeptide, without any presumptions as to the conformation of the underlying primary or secondary structure. The method involves computer simulation of the polypeptide, and more particularly simulating a real-size primary structure in an aqueous environment, shrinking the size of the polypeptide isobarically and isothermally, and expanding the simulated polypeptide to its real size in selected time periods. A useful set of tools, termed Balaji plots, energy conformational maps, and probability maps, assist in identifying those portions of the predicted peptide structure that are most flexible or most rigid. The rational design of novel compounds, useful as drugs, e.g., bioactive peptidomimetic compounds, and constrained analogs thereof, is thus made possible using the simulation methods and tools of the described invention.
CLM What is claimed is:
1. An ab initio computer-assisted method of predicting a stable tertiary structure of a **peptide** without any presumption regarding the underlying structural characteristics of the **peptide**, comprising: (a) simulating a real-size primary structure of a **polypeptide** in a solvent box, said primary structure comprising a plurality of amino acid residues linked together in a chain, each residue having .phi., .psi. angles associated therewith, said .phi., .psi. angles defining the relative angle of a first and second amide plane of said amino acid residue with a common C.sup..alpha. atom of said amino acid residue; (b) shrinking the size of the **peptide** isobarically and isothermally; (c) expanding the **peptide** to its real size in selected time periods; and (d) measuring the .phi., .psi. angles of the expanding amino acid residues.

2. The method of predicting a tertiary structure as set forth in claim 1 wherein step (d) further includes measuring the energy state of the expanding residues.

3. The method of predicting a tertiary structure as set forth in claim 1 wherein step (c) further includes expanding the **peptide** beyond

its real size in selected time periods.

4. The method of predicting a tertiary structure as set forth in claim 3 further including analyzing the .phi., .psi. angles corresponding to at least two consecutive selected time periods in order to identify the differences therebetween, said differences being indicative of the rigidity of a particular amino acid residue within said **polypeptide** chain.

5. The method of predicting a tertiary structure as set forth in claim 4 wherein said step of analyzing the .phi., .psi. angles includes plotting the .phi., .psi. angles of the simulated **peptide** as a function of the residue.

6. The method of predicting a tertiary structure as set forth in claim 5 wherein the step of plotting the .phi., .psi. angles comprises (i) plotting the .phi. angle as the base of a wedge and the .psi. angle as the tip of a wedge along a first axis of a plot, said first axis having angular values marked thereon, with the tip of the wedge being aligned with the value of the .psi. angle as indicated on the first axis, and the base of the wedge being aligned with the value of the .phi. angle as indicated on the first axis, and (ii) plotting a separate wedge for each amino acid residue along a second axis of said plot, said second axis being orthogonal to said first axis, said second axis having numerical values marked thereon, the wedge for a particular residue being aligned with the number of the residue as indicated on the second axis, whereby a wedge appears on said plot for each amino acid residue in said **polypeptide** chain, with the location of the base and tip of each wedge relative to said first axis indicating the .phi., .psi. angles, respectively, for the particular amino acid residue indicated by the location of the wedge relative to said second axis.

7. The method of predicting a tertiary structure as set forth in claim 1 wherein step (c) comprises expanding the amino acid residues of said **polypeptide** chain one at a time.

8. The method of predicting a tertiary structure as set forth in claim 7 further including biasing the expansion towards a structure predicted by known chemical and physical data.

9. The method of predicting a tertiary structure as set forth in claim 1 wherein step (c) comprises expanding the amino acid residues of said **polypeptide** chain simultaneously.

10. The method of predicting a tertiary structure as set forth in claim 9 further including biasing the expansion towards a structure predicted by known chemical and physical data.

11. A computer-assisted method for determining areas of flexibility and rigidness in a **peptide**, said **peptide** comprising a plurality of residues linked together in a chain, said method comprising the steps of: (a) electronically simulating said **peptide** in a fluid environment where the residue chain is free to move and fold as a result of natural molecular or electrical forces present in said residues; (b) measuring the .phi., .psi. angles associated with each residue of said simulated **peptide** at discrete time periods as said residue chain moves in said environment; (c) plotting the .phi., .psi. angles of the stimulated **peptide** as a function of the residue for a plurality of consecutive discrete time periods; and (d) determining the differences between the .phi., .psi. angles of corresponding residues of adjacent discrete time periods, whereby the relative flexibility or rigidness of a particular bond within said

peptide is identified.

12. The method for determining areas of flexibility and rigidness in a **peptide** as set forth in claim 11 wherein step (a) includes: (i) shrinking the size of the simulated **peptide** isobarically and isothermally while in said fluid environment, and (ii) expanding the simulated **peptide** to its real size in discrete steps at each of said discrete time periods.

13. The method for determining areas of flexibility and rigidness in a **peptide** as set forth in claim 12 further including expanding the simulated **peptide** beyond its real size in discrete time periods.

14. The method for determining areas of flexibility and rigidness in a **peptide** as set forth in claim 13 wherein step (c) comprises (i) plotting the .phi. angle as the base of a wedge and the .psi. angle as the tip of a wedge along a first axis of a plot, said first axis having angular values marked thereon, with the tip of the wedge being aligned with the value of the .psi. angle as indicated on the first axis, and the base of the wedge being aligned with the value of the .phi. angle as indicated on the first axis, and (ii) plotting a separate wedge for each amino acid residue along a second axis of said plot, said second axis being orthogonal to said first axis, said second axis having numerical values marked thereon, the wedge for a particular residue being aligned with the number of the residue as indicated on the second axis, whereby a wedge appears on said plot for each amino acid residue in said **polypeptide** chain, with the location of the base and tip of each wedge relative to said first axis indicating the .phi., .psi. angles, respectively, for the particular amino acid residue indicated by the location of the wedge relative to said second axis.

15. A system for determining areas of flexibility and rigidness in a **peptide**, said **peptide** comprising a plurality of residues linked together in a chain, said system comprising: (a) simulating means for simulating the structure of said **peptide** in a fluid environment where the residue chain is free to move and fold as a result of natural molecular or electrical forces present in said residues; (b) recording means for recording the .phi., .psi. angles associated with each residue of said simulated **peptide** at discrete time periods as said residue chain moves in said environment; (c) plotting means for plotting the .phi., .psi. angles of the simulated **peptide** as a function of the residue for a plurality of consecutive discrete time periods; and (d) analyzing means for analyzing the differences between the .phi., .psi. angles of corresponding residues of adjacent discrete time periods to identify the relative flexibility or rigidness of a particular bond within said **peptide**.

16. The system for determining areas of flexibility and rigidness in a **peptide** as set forth in claim 15 wherein said simulation means includes processing means for shrinking the size of the simulated **peptide** isobarically and isothermally while in said fluid environment, and expanding the simulated **peptide** to its real size in discrete steps at each of said discrete time periods.

17. The system for determining areas of flexibility and rigidness in a **peptide** as set forth in claim 16 wherein said processing means is for further expanding the simulated **peptide** beyond its real size in discrete time periods.

18. The system for determining areas of flexibility and rigidness in a

peptide as set forth in claim 17 wherein said plotting means includes: means for plotting the .phi. angle as the base of a wedge and the .psi. angle as the tip of a wedge along a first axis of a plot, said first axis having angular values marked thereon, with the tip of the wedge being aligned with the value of the .psi. angle as indicated on the first axis, and the base of the wedge being aligned with the value of the .phi. angle as indicated on the first axis, and means for plotting a separate wedge for each amino acid residue along a second axis of said plot, said second axis being orthogonal to said first axis, said second axis having numerical values marked thereon, the wedge for a particular residue being aligned with the number of the residue as indicated on the second axis, whereby a wedge appears on said plot for each amino acid residue in said **polypeptide** chain, with the location of the base and tip of each wedge relative to said first axis indicating the .phi., .psi. angles, respectively, for the particular amino acid residue indicated by the location of the wedge relative to said second axis.

19. A method for generating biologically or pharmacologically active molecules, said method comprising: (i) determining the amino acid sequence of the hypervariable region of a monoclonal antibody having biological or pharmacological activity, and (ii) producing a peptidomimetic compound based on the amino acid sequence of step (i), wherein said peptidomimetic compound substantially retains the biological or pharmacological activity of said monoclonal antibody.

L3 ANSWER 66 OF 73 USPATFULL
AN 96:109612 USPATFULL
PI US 5579250 19961126
PI US 5579250 19961126
TI Method of rational drug design based on AB initio computer simulation of conformational features of **peptides**
AB A method of rational drug design includes simulating polypeptides in a way that predicts the most probable secondary and/or tertiary structures of a polypeptide, e.g., an oligopeptide, without any presumptions as to the conformation of the underlying primary or secondary structure. The method involves computer simulation of the polypeptide, and more particularly simulating a real-size primary structure in an aqueous environment, shrinking the size of the polypeptide isobarically and isothermally, and expanding the simulated polypeptide to its real size in selected time periods. A useful set of tools, termed Balaji plots, energy conformational maps, and probability maps, assist in identifying those portions of the predicted peptide structure that are most flexible or most rigid. The rational design of novel compounds, useful as drugs, e.g., bioactive peptidomimetic compounds, and constrained analogs thereof, is thus made possible using the simulation methods and tools of the described invention.
CLM What is claimed is:
1. A computer-assisted method of rational design of bioactive compounds, comprising: (a) electronically simulating and selecting the most probable conformations of a given **polypeptide**, wherein the most probable conformation of a given **polypeptide** is selected by an ab initio procedure, comprising: (i) simulating a real-size primary structure of a **polypeptide** in a solvent box, wherein the **polypeptide** comprises a plurality of amino acid residues linked together in a chain, each residue having .phi., .psi. angles associated therewith, said .phi., .psi. angles defining the relative angle of a first and second amide plane of said amino acid residue with a common C.sup..alpha. atom of said amino acid residue; (ii) shrinking the size of the **polypeptide** isobarically and isothermally; and (iii) expanding the **polypeptide** to its real size in selected

time periods to determine an energetically most probable three-dimensional structure of the **peptide**; (b) designing a chemically modified analog that substantially mimics the energetically most probable three-dimensional structure of the **peptide**; (c) chemically synthesizing the chemically modified analog of the **peptide**; and (d) evaluating the bioactivity of the synthesized chemically modified analog and selecting the analogs that exhibit bioactivity.

2. The method of rational design of claim 1, further comprising: (e) designing a peptidomimetic based on the conformation of the synthesized chemically modified analog.

3. The method of rational design of claim 1, wherein step (a) further comprises: (iv) recording the .phi., .psi. angles of the amino acid residues of the expanded **peptide**.

4. The method of rational design of claim 3 wherein step (a) further comprises: (v) determining the energy state of the expanded amino acid residues.

5. The method of rational design of claim 4, further comprising expanding the **polypeptide** beyond its real size in selected time increments, and recording the .phi., .psi. angles of the residues thus expanded.

6. The method of rational design of claim 4, wherein the step of simulating and selecting the most probable conformation of the **peptides** includes analyzing the recorded .phi., .psi. angles and energy states for each residue as a function of the time increments.

7. The method of rational design of claim 6, wherein the step of analyzing the recorded .phi., .psi. angles comprises plotting the .phi., .psi. angles for each amino acid residue of the simulated **polypeptide** as a function of the location of the residue within the **polypeptide** chain.

8. The method of rational design of claim 7 wherein the step of plotting the .phi., .psi. angles comprises: (i) plotting the .phi. angle as the base of a wedge and the .psi. angle as the tip of a wedge along a first axis of a plot, the first axis having angular values marked thereon, with the top of the wedge being aligned with the value of the .psi. angle as indicated on the first axis, and the base of the wedge being aligned with the value of the angle as indicated on the first axis, and (ii) plotting a separate wedge for each amino acid residue along a second axis of the plot, the second axis being orthogonal to the first axis, the second axis having numerical values marked thereon, the wedge for a particular residue being aligned with the number of the residue as indicated on the second axis, whereby a wedge appears on the plot for each amino acid residue in the **polypeptide** chain, with the location of he base and tip of each wedge relative to the first axis indicating the .phi., .psi. angles, respectively, for the particular amino acid residue indicated by the location of the wedge relative to the second axis.

9. The method of rational design of claim 6, further including analyzing the differences between the .phi., .psi. angles of corresponding residues at selected time increments to identify the relative flexibility or rigidness of a particular bond within the **polypeptide**.

10. The method of rational design of claim 9 wherein step (b) of

designing and synthesizing a chemically modified analog of the selected **peptide** comprises identifying flexible portions of the **polypeptide** chain and replacing the flexible portions with bioisostere moieties.

11. The method of claim 10, wherein the flexible portions are identified by: (i) electronically simulating the **peptide** in a fluid environment where the residue chain is free to move and fold as a result of natural molecular or electrical forces present in the residues; (ii) measuring the .phi., .psi. angles associated with each residue of the simulated **peptide** at discrete time periods as the residue chain moves in the environment; (iii) plotting the .phi.,.psi. angles of the simulated **peptide** as a function of the residue for a plurality of consecutive discrete time periods; and (iv) determining the differences between the .phi., .psi. angles of corresponding residues of adjacent discrete time periods, whereby the relative flexibility or rigidness of a particular bond within the **peptide** is identified.

12. The method of claim 11, wherein step (i) comprises: (A) shrinking the size of the simulated **peptide** isobarically and isothermally while in the fluid environment; and (B) expanding the simulated **peptide** to its real size in discrete steps at each of the discrete time periods.

13. The method of claim 12, wherein the steps for determining areas of flexibility and rigidness in a **peptide** further comprise: (C) expanding the simulated **peptide** beyond its real size in discrete time periods.

14. The method of claim 13, wherein the step (iii) comprises: (A) plotting the .phi. angle as the base of a wedge and the .psi. angle as the tip of a wedge along a first axis of a plot, the first axis having angular values marked thereon, with the tip of the wedge being aligned with the value of the .psi. angle as indicated on the first axis, and the base of the wedge being aligned with the value of the .phi. angle as indicated on the first axis; and (B) plotting a separate wedge for each amino acid residue along a second axis of the plot, the second axis being orthogonal to the first axis, the second axis having numerical values marked thereon, the wedge for a particular residue being aligned with the number of the residue as indicated on the second axis, whereby a wedge appears on the plot for each amino acid residue in the **polypeptide** chain, with the location of the base and tip of each wedge relative to the first axis indicating the .phi., .psi. angles, respectively, for the particular amino acid residue indicated by the location for the wedge relative to the second axis.

15. The method of claim 14, wherein step (d) further comprises measuring the energy state of the expanding residues.

16. The method of claim 14, wherein step (c) further includes expanding the **peptide** beyond its real size in selected time periods.

17. The method of claim 16, further comprising analyzing the .phi., .psi. angles corresponding to at least two consecutive selected time periods in order to identify the differences therebetween, wherein the differences are indicative of the rigidity of a particular amino acid residue within the **polypeptide** chain.

18. The method of claim 17, wherein the step of analyzing the .phi., .psi. angles includes plotting the .phi., .psi. angles of the simulated **peptide** as a function of the residue.

19. The method of claim 18, wherein the step of plotting the .phi., .psi. angles comprises: (i) plotting the angle as the base of a wedge and the .phi. angle as the tip of a wedge along a first axis of a plot, the first axis having angular values marked thereon, with the tip of the wedge being aligned with the value of the .psi. angle as indicated on the first axis, and the base of the wedge being aligned with the value of the .phi. angle as indicated on the first axis, and (ii) plotting a separate wedge for each amino acid residue along a second axis of the plot, the second axis being orthogonal to the first axis, the second axis having numerical values marked thereon, the wedge for a particular residue being aligned with the number of the residue as indicated on the second axis, whereby a wedge appears on the plot for each amino acid residue n the **polypeptide** chain, with the location of the base and tip of each wedge relative to the first axis indicating the .phi., .psi. angles, respectively, for the particular amino acid residue indicated by the location of the wedge relative to the second axis.

20. The method of claim 14, wherein step (c) comprises expanding the amino acid residues of the **polypeptide** chain one at a time.

21. The method of claim 20, further comprising biasing the expansion towards a structure predicted by known chemical and physical data.

22. The method of claim 14, wherein step (c) comprises expanding the amino acid residues of the **polypeptide** chain simultaneously.

23. The method of claim 22, further comprising biasing the expansion towards a structure predicted by known chemical and physical data.

24. A method for the design of a peptidomimetic or pharmacophore, the method comprising: (1) determining the energetically most probable tertiary structure of that portion of a pharmaceutically active compound that is responsible for the pharmacological action of the compound (2) producing a simulated, chemically modified **peptide** or peptidomimetic structure that substantially mimics the energetically most probable three-dimensional structure of the pharmaceutically active compound; (3) chemically synthesizing the chemically modified **peptide** or peptidomimetic structure; and (4) evaluating the bioactivity of the synthesized **peptide** or peptidomimetic structure.

25. The method of claim 24, wherein producing a simulated, chemically modified **peptide** or peptidomimetic structure that substantially mimics the energetically most probable three-dimensional structure of the pharmaceutically active compound is carried out by: (i) determining the .phi. and .psi. angles for each residue included in the energetically most probable conformation of the pharmaceutically active compound; (ii) comparing the .phi. and .psi. angles for each residue obtained in step (i) with the .phi. and .psi. angles for each residue of known **polypeptide** species, and (iii) substituting a chemically modified moiety for at least one of the residues of the pharmaceutically active compound, wherein the chemically modified moiety has .phi. and .psi. angles that are substantially similar to the .phi. and .psi. angles of the residue to be replaced.

26. The method of claim 24, wherein the energetically most probable tertiary structure of that portion of a pharmaceutically active compound which is responsible for the pharmacological action is determined by the method, comprising: (a) simulating a real-size primary structure of a **polypeptide** in a solvent box, the primary structure comprising a plurality of amino acid residues linked together in a chain, each

residue having .phi., .psi. angles associated therewith, the .phi., .psi. angles defining the relative angle of a first and second amide plane of the amino acid residue with a common C.sup..alpha. atom of the amino acid residue; (b) shrinking the size of the **peptide** isobarically and isothermally; (c) expanding the **peptide** to its real size in selected time periods; and (d) measuring the .phi., .psi. angles of the expanding amino acid residues.

L3 ANSWER 70 OF 73 USPATFULL
AN 93:72614 USPATFULL
PI US 5241470 19930831
PI US 5241470 19930831
TI Prediction of **protein** side-chain conformation by packing optimization
AB A method is provided for determining the packing conformation of amino acid side chains on a fixed peptide backbone. Using a steric interaction potential, the side chain atoms are rotated about carbon-carbon bonds such that the side chains preferably settle in a low energy packing conformation. Rotational moves are continued according to a simulated annealing procedure until a set of low energy conformations are identified. These conformations represent the structure of the actual peptide. The method may be employed to identify the packing configuration of mutant peptides.
CLM What is claimed is:
1. A method for determining the three dimensional structure of a **peptide**, the **peptide** having amino acid side chains extending from a defined main chain backbone, each amino acid side chain having predefined rotational degrees of freedom, the method comprising the steps of: a. inputting coordinates of the main chain backbone of said **peptide**; b. constructing an initial three dimensional **peptide** conformation by placing the amino acid side chains on said main chain backbone coordinates, the **peptide** being in an initial three dimensional **peptide** conformation; c. randomly rotating said amino acid side chains around said predefined rotational degrees of freedom by small rotational perturbations to produce a modified three dimensional **peptide** conformation; d. determining the side chain steric interaction energy for said modified **peptide** conformation; e. creating a final three dimensional **peptide** conformation by reducing said side chain steric interaction energy by repeating steps c-d, wherein said step of randomly rotating is biased toward conformations having lower values of said side chain steric interaction energy and wherein said interaction energy is truncated if it exceeds a preselected maximum.

2. A method according to claim 1 wherein step e is conducted with simulated annealing.

3. A method according to claim 2 further comprising the steps of: e. creating additional final three dimensional **peptide** conformations by conducting steps b-e repeatedly; and f. averaging said final three dimensional **peptide** conformations to produce an average three dimensional **peptide** conformation.

4. A method according to claim 2 wherein the step of reducing said side chain steric interaction energy comprises minimizing said side chain steric interaction energy of the **peptide**.

5. A method according to claim 3 wherein the step of averaging said three dimensional **peptide** conformations comprises the step of selecting an energetically stable side chain conformation for each side chain of the **peptide**, wherein each said energetically stable

side chain conformation is selected from the group consisting of corresponding side chain conformations from said three dimensional models.

6. A method according to claim 5 wherein each selected energetically stable side chain conformation has the lowest steric interaction energy.

7. A method according to claim 1 wherein the defined main chain backbone of said **peptide** comprises C.sub.i.sup..alpha., N.sub.i, I.sub.i, and C.sub.i of each said amino acid.

8. A method according to claim 1 wherein the step of constructing an initial three-dimensional **peptide** conformation comprises: a. determining the three dimensional position of C.sub.i.sup..alpha. for each amino acid side chain; and b. assigning a torsion angle to each predefined rotational degree of freedom.

9. A method according to claim 8 wherein each said torsion angle is selected randomly.

10. A method according to claim 8 wherein a plurality of torsion angles are selected randomly and a plurality of torsion angles are predefined.

11. A method according to claim 1 wherein said steric interaction energy is calculated according to the Lennard-Jones potential: ##EQU1## wherein r is the interatomic distance; r.sub.0 is the equilibrium interatomic distance; and .epsilon..sub.0 is the depth of energy well for the interaction.

12. A method according to claim 11 wherein the Lennard-Jones potential is truncated to a predetermined maximum energy in the range of about 4 to 15 kcal/mol.

13. A method according to claim 1 further comprising a step of determining torsional interaction energies between adjacent carbon atoms.

14. A method according to claim 13 wherein said torsional interaction energy is calculated according to the equation: E.sub.torsion =K cos [n(.chi.-d)] wherein K is an empirical energy constant, n and d are constants and .chi. is a torsion angle between adjacent carbon atoms.

15. A method according to claim 14 where K is between about 1 and about 5 Kcal/mol, and wherein n is 3 and d is 0.

16. A method according to claim 1 wherein each amino acid side chain rotational degree of freedom is rotated by an angle randomly selected in the range between -25.degree. and 25.degree. .

17. A method according to claim 16 wherein each amino acid side chain rotational degree of freedom is rotated by an angle randomly selected in the range between -12.degree. and 12.degree..

18. A method according to claim 1 wherein each amino acid side chain rotational degree of freedom is rotated by an angle randomly selected from the group consisting of approximately -10.degree., approximately 0.degree., and approximately 10.degree..

19. The method according to claim 1 wherein the step of randomly rotating is biased toward conformations having lower values of said side chain steric interaction energy by selectively accepting said modified three dimensional **peptide** conformations, the step of

selectively accepting comprising: comparing the steric interaction energy of a current modified **peptide** conformation with the interaction energy of a previous **peptide** conformation; and reverting to said previous **peptide** conformation according to a predetermined probability when the interaction energy of said modified **peptide** conformation is higher than the interaction energy of said previous **peptide** conformation.

20. A method according to claim 19 wherein said predetermined probability is represented by: P=exp (-E.sub.diff /kT), wherein E.sub.diff is the steric interaction energy difference between the modified **peptide** conformation and the previous **peptide** conformation, k is the boltzman constant, and T is a predetermined constant.

21. A method for determining the three dimensional structure of a **peptide**, the **peptide** having a plurality of amino acid side chains extending from a defined main chain backbone, each amino acid side chain having predefined rotational degrees of freedom, and the plurality of side chains having a plurality of conformations defining a conformation space, the method comprising the steps of: a. constructing an initial three dimensional **peptide** conformation by placing each amino acid side chain in an initial three dimensional conformation; b. determining a side chain steric interaction energy for said initial **peptide** conformation; and c. searching the full conformation space for low energy **peptide** conformations by randomly rotating each of said plurality of amino acid side chains around respective predefined rotational degrees of freedom to produce a modified three dimensional **peptide** conformation and determining side chain steric interaction energy for said modified **peptide** conformation, said low energy **peptide** conformations representing the three dimensional structure of said **peptide**.

22. A method of producing a three-dimensional image of a **peptide** with the aid of a digital computer, the **peptide** having a primary sequence, main chain coordinates, and side chains bonded to the main chain, the side chains comprising atoms connected to one another and the main chain by side chain bonds, the method comprising the following steps: storing the primary sequence, the main chain coordinates, and the side chains in a computer useable form; repeatedly moving selected side chain atoms by rotation about selected side chain bonds to conformations having a low steric interaction potential, the rotation distance and direction determined by simulated annealing; producing a final three dimensional conformation of the **peptide** by conducting simulated annealing for a predetermined length; and displaying an image of the final three dimensional conformation on a display monitor.

23. The method recited in claim 22 further comprising a step of storing conformations having steric interaction potentials below a predefined value.

24. The method recited in claim 22 wherein the steric interaction potential is determined according to the Lennard-Jones potential, and wherein the value of the Lennard-Jones potential is truncated when it exceeds a predetermined value.

25. A system for determining the three-dimensional conformation of a **peptide**, the **peptide** having a primary sequence, main chain coordinates, and side chains bonded to the main chain, the side chains comprising atoms connected to one another and the main chain by

side chain bonds, the system comprising: means for converting the primary sequence and main chain coordinates of the **peptide** to a computer useable form; means for bonding the side chains to the main chain in a random orientation to form an initial **peptide** conformation; means for rotating selected side chain atoms about selected side chain bonds to form intermediate **peptide** conformations; means for determining the steric interaction energy of the intermediate **peptide** conformations; means for condensing the intermediate **peptide** conformations to produce a final **peptide** conformation by simulated annealing; and means for displaying images of the final **peptide** conformation.

26. The system of claim 25 wherein means for displaying images is a computer display terminal.

L3 ANSWER 71 OF 73 USPATFULL
AN 90:75530 USPATFULL
PI US 4959796 19900925
PI US 4959796 19900925
TI Method of producing analytical curve
AB A method of producing an analytical curve for an analyzing apparatus which provides an analysis result on the basis of the analytical curve in response to a measurement value obtained by photoelectrically measuring light intensity reflected from a slide to be analyzed. A plurality of reference slides are measured by first analyzing apparatus which has a predetermined analytical curve, thereby obtaining a plurality of first measurement values and providing a plurality of first analysis results. The plurality of reference slides are further measured by second analyzing apparatus, thereby obtaining a plurality of second measurement values. Analytical curve for the second analyzing apparatus for produced on the basis of a relation between the first measurement values and the second measurement values so that a plurality of second analysis results correspond to the plurality of first analysis results.
CLM What is claimed is:
1. A method of producing an analytical curve for an analyzing apparatus, the apparatus providing an analysis result on the basis of said analytical curve in response to a measurement value obtained by photoelectrically measuring light intensity reflected from a slide to be analyzed, the method comprising the steps of: measuring a plurality of reference slides by using a first analyzing apparatus which has a predetermined analytical curve, thereby obtaining a plurality of first measurement values and providing a plurality of first analysis results; measuring said plurality of reference slides by using a second analyzing apparatus, thereby obtaining a plurality of second measurement values; and producing an analytical curve for said second analyzing apparatus to provide a plurality of second analysis results on the basis of a relation between said first measurement values and said second measurement values so that said plurality of second analysis results correspond to said plurality of first analysis results.

2. The method of claim 1, wherein said reference slide is made of color paper, plastic or ceramic.

3. The method of claim 1, wherein said reference slide is dyed in a predetermined color so as to provide a predetermined reflection density.

4. The method of claim 1, wherein said measuring step includes the step of storing said measurement values in a memory.

5. The method of claim 1, wherein said predetermined analytical curve is represented by a conversion formula having a variable measurement value,

and wherein the step of producing an analytical curve comprises the substeps of processing said first measurement values and said second measurement values so as to provide a regression formula, substituting said variable measurement value with said regression formula to provide a revised conversion formula, and storing the revised conversion formula as said analytical curve for said second analyzing apparatus in a memory of said second analyzing apparatus.

6. A method of claim 1, wherein said relation between said first measurement values and said second measurement values is represented by a regression formula and said predetermined analytical curve is represented by a conversion formula, and wherein said analytical curve for said second analyzing apparatus is represented by a formula obtained by combining said regression formula and said conversion formula.

7. The method of claim 5, wherein said conversion formula is represented by $Y=B/(X-A)+C$, in which Y is an analysis result, X is a measurement value, and A, B and C are constants.

8. The method of claim 6, wherein said regression formula is a primary regression formula represented by $X=aX'+b$ in which X is a measurement value obtained by the first analyzing apparatus, X' is a measurement value obtained by the second analyzing apparatus, and a and b are constant.

9. The method of claim 1, wherein said first and second analyzing apparatus provide an analysis result for analysis items using End-Point Assay and Rate Assay.

10. The method of claim 9, wherein said analysis items in accordance with End-Point Assay are glucose, total cholesterol, hemoglobin, urea nitrogen, urea acid, total **protein**, albumin, triglyceride, and total bilirubin.

11. The method of claim 9, wherein said analysis items analyzed using Rate Assay are glutamic-oxaloacetic transaminase, glutamin-pyruvic transaminase, alkaline phosphatase, and lactate dehydrogenase.

12. The method of claim 1, wherein said step of measuring a plurality of reference slides by both said first and second analyzing apparatus comprises substeps of loading said plurality of reference slides onto said analyzing apparatus, checking the condition of light source by a calibration mechanism, and performing a photometric operation for said loaded reference slides with said checked light source.

13. The method of claim 12, wherein said step of checking the calibration of light source includes the steps of calibrating wavelength, judging the sufficiency of light intensity, and judging whether the end of the service life has been reached.

```
=> act s788006/l
L1  (     21974)SEA FILE=CAPLUS ABB=ON  PLU=ON  ALPHA (W)(HELIX OR HELIC##)
L2  (      8921)SEA FILE=CAPLUS ABB=ON  PLU=ON  L1 AND (IDENTIF? OR DETERMIN? O
L3  (      2208)SEA FILE=CAPLUS ABB=ON  PLU=ON  L2 AND MODEL?
L4  (        41)SEA FILE=CAPLUS ABB=ON  PLU=ON  (((ALPHA/IT (W)(HELIX/IT OR HEL
L5  (        41)SEA FILE=CAPLUS ABB=ON  PLU=ON  L3 AND L4
L6  (       337)SEA FILE=CAPLUS ABB=ON  PLU=ON  ((ALPHA/IT (W)(HELIX/IT OR HELI
L7  (      1217)SEA FILE=CAPLUS ABB=ON  PLU=ON  (COILED COIL)/IT
L8  (      1540)SEA FILE=CAPLUS ABB=ON  PLU=ON  L6 OR L7
L9  (       163)SEA FILE=CAPLUS ABB=ON  PLU=ON  L8 AND MODEL?/IT
L10 (        67)SEA FILE=CAPLUS ABB=ON  PLU=ON  L9 AND L2

=> s l10
        394553 ALPHA/IT
            24 ALPHAS/IT
        394558 ALPHA/IT
                 ((ALPHA OR ALPHAS)/IT)
         14462 HELIX/IT
          1159 HELIXES/IT
           122 HELICES/IT
         15222 HELIX/IT
                 ((HELIX OR HELIXES OR HELICES)/IT)
         12849 HELIC##/IT
          6031 ALPHA/IT (W)(HELIX/IT OR HELIC##/IT)
         72579 IDENTIF?/IT
         29445 DETERMIN?/IT
          4124 DET/IT
           705 DETS/IT
          4821 DET/IT
                 ((DET OR DETS)/IT)
         15654 DETD/IT
         13213 DETG/IT
        801570 DETN/IT
          5732 DETNS/IT
        803942 DETN/IT
                 ((DETN OR DETNS)/IT)
        855095 DETERMIN?/IT
                 ((DETERMIN? OR DET OR DETD OR DETG OR DETN)/IT)
         14726 LOCAT?/IT
          2655 COILED/IT
          8059 COIL/IT
          6173 COILS/IT
         12784 COIL/IT
                 ((COIL OR COILS)/IT)
          1217 (COILED COIL)/IT
                 ((COILED(W)COIL)/IT)
        561125 MODEL?/IT
       1279200 ALPHA
          2462 ALPHAS
       1279290 ALPHA
                 (ALPHA OR ALPHAS)
         47926 HELIX
         11540 HELIXES
          3213 HELICES
         55194 HELIX
                 (HELIX OR HELIXES OR HELICES)
         47814 HELIC##
         21979 ALPHA (W)(HELIX OR HELIC##)
        744856 IDENTIF?
        716617 DETERMIN?
        499871 DET
```

```
         30293 DETS
        527864 DET
                  (DET OR DETS)
       1756953 DETD
             2 DETDS
       1756955 DETD
                  (DETD OR DETDS)
        281249 DETG
       1181860 DETN
        123115 DETNS
       1263399 DETN
                  (DETN OR DETNS)
       3424156 DETERMIN?
                  (DETERMIN? OR DET OR DETD OR DETG OR DETN)
        334375 LOCAT?
L11         67 L9 AND L2

=> d au,so,kwic 2,10,17,18,47,52

L11  ANSWER 2 OF 67  CAPLUS  COPYRIGHT 2002 ACS
AU   McClain, Diana L.; Binfet, Joseph P.; Oakley, Martha G.
SO   Journal of Molecular Biology (2001), 313(2), 371-383
     CODEN: JMOBAK; ISSN: 0022-2836
AB   Coiled coils are formed by two or more .alpha.-helixes
     that align in a parallel or an antiparallel relative orientation.  The
     factors that det. a preference for a given relative helix
     orientation are incompletely understood.  The helix orientation preference
     for the designed coiled coil, Acid-a1-Base-a1, was measured previously.
     This model system therefore provides a means for the exptl. detn
     . of the energetic contribution of a variety of interactions to helix
     orientation specificity.  The antiparallel preference for Acid-a1-Base-a1
     is imparted.  .  .  proposed to influence helix orientation preference.
     In the Acid-a1-Base-a1 heterodimer, potentially attractive Coulombic
     interactions are expected in both orientations.  To det. the
     energetic consequences of Coulombic interactions for helix orientation
     preference, we have positioned a single charged residue in each peptide.
     .  .
IT   Coiled-coil
     Conformational free energy
     Electrostatic force
     Molecular orientation
     Protein folding
        (evaluation of the energetic contribution of interhelical Coulombic
        interactions for coiled coil helix orientation
        specificity)
IT   Molecular association
        (heterodimerization; evaluation of the energetic contribution of
        interhelical Coulombic interactions for coiled coil
        helix orientation specificity)
IT   Conformation
        (protein; evaluation of the energetic contribution of interhelical
        Coulombic interactions for coiled coil helix
        orientation specificity)
IT   386769-16-0   386769-17-1   386769-18-2   386769-19-3   386769-20-6
     RL: BSU (Biological study, unclassified); PRP (Properties); BIOL
     (Biological study)
        (model peptide; evaluation of the energetic contribution of
        interhelical Coulombic interactions for coiled coil
        helix orientation specificity)

L11  ANSWER 10 OF 67  CAPLUS  COPYRIGHT 2002 ACS
AU   Newman, John R. S.; Wolf, Ethan; Kim, Peter S.
```

SO Proceedings of the National Academy of Sciences of the United States of America (2000), 97(24), 13203-13208
CODEN: PNASA6; ISSN: 0027-8424

TI A computationally directed screen **identifying** interacting coiled coils from Saccharomyces cerevisiae

AB Computational methods can frequently **identify** protein-interaction motifs in otherwise uncharacterized open reading frames. However, the **identification** of candidate ligands for these motifs (e.g., so that partnering can be **detd.** exptl. in a directed manner) is often beyond the scope of current computational capabilities. One exception is provided by the coiled-coil interaction motif, which consists of two or more **.alpha. helixes** that wrap around each other: the ligands for coiled-coil sequences are generally other coiled-coil sequences, thereby greatly simplifying the motif/ligand recognition problem. Here, we describe a two-step approach to **identifying** protein-protein interactions mediated by two-stranded coiled coils that occur in Saccharomyces cerevisiae. Coiled coils from the yeast genome are first predicted computationally, by using the MULTICOIL program, and assocns. between coiled coils are then **detd.** exptl. by using the yeast two-hybrid assay. We report 213 unique interactions between 162 putative coiled-coil sequences. We evaluate the resulting interactions, focusing on assocns. **identified** between components of the spindle pole body (the yeast centrosome).

IT Computer program
(MULTICOIL; a computationally directed screen **identifying** interacting **coiled coils** from Saccharomyces cerevisiae)

IT Saccharomyces cerevisiae
Simulation and **Modeling**, physicochemical
(a computationally directed screen **identifying** interacting **coiled coils** from Saccharomyces cerevisiae)

IT **Coiled-coil**
Conformation
(protein; a computationally directed screen **identifying** interacting **coiled coils** from Saccharomyces cerevisiae)

IT Genetic methods
(yeast two-hybrid assay; a computationally directed screen **identifying** interacting **coiled coils** from Saccharomyces cerevisiae)

L11 ANSWER 17 OF 67 CAPLUS COPYRIGHT 2002 ACS

AU Bourla, Lisa; Seifer, Tidhar; Honig, Barry; Ben-Tal, Nir

SO Frontiers Science Series (2000), 30(Currents in Computational Molecular Biology), 157-158
CODEN: FCFUEO; ISSN: 0915-8502

TI The **identification** of transmembrane helices in the sequences of membrane proteins using a computationally-derived hydrophobicity scale

AB . . . of a computationally-derived hydrophobicity scale for the transfer of amino acids from water to bilayers in the context of an . **alpha.-helix** is described. Continuum solvent models have been used to calc. the transfer free energies of polyalanine . **alpha.-helixes** from the aq. phase into lipid bilayers and the results were in very good agreement with exptl. data. In this. . . energies of the 20 amino acids from the aq. phase into the lipid bilayer in the context of a polyalanine **.alpha.-helix.** The scale in a dynamic programming algorithm was then used to **identify** transmembrane spans in the sequences of membrane proteins. The algorithm is based on a summation of the free energies of. . . was tested on a set of over 140 bacterial and eukaryotic integral membrane protein sequences. The transmembrane spans were correctly

identified in about 60% of the proteins in the set. Comparison of current results with results from predictions which used the. . .

IT Membrane, biological
(bilayer; **identification** of transmembrane helixes in sequences of membrane proteins using a computationally-derived hydrophobicity scale)

IT Hydrophobicity
.alpha.-Helix
(**identification** of transmembrane helixes in sequences of membrane proteins using a computationally-derived hydrophobicity scale)

IT Proteins, specific or class
RL: PRP (Properties)
(membrane, integral; **identification** of transmembrane helixes in sequences of membrane proteins using a computationally-derived hydrophobicity scale)

IT Conformation
(protein; **identification** of transmembrane helixes in sequences of membrane proteins using a computationally-derived hydrophobicity scale)

IT Free energy of transfer
(use of continuum solvent **models** to calc. the transfer free energies of polyalanine **.alpha.-helixes** from the aq. phase into lipid bilayers)

L11 ANSWER 18 OF 67 CAPLUS COPYRIGHT 2002 ACS
AU Sun, Jia Ke; Penel, Simon; Doig, Andrew J.
SO Protein Science (2000), 9(4), 750-754
CODEN: PRCIEI; ISSN: 0961-8368
TI **Determination** of **.alpha.-helix** N1 energies after addition of N1, N2, and N3 preferences to helix/coil theory
AB . . . that amino acids show unique structural preferences for the N1, N2, and N3 positions in the first turn of the **.alpha.-helix**. We have therefore extended helix-coil theory to include statistical wts. for these **locations**. The helix content of a peptide in this model is a function of N-cap, C-cap, N1, N2, N3, C1, and. . .

IT Conformational free energy
.alpha.-Helix
(**detn**. of **.alpha.-helix** N1 energies after addn. of N1, N2, and N3 prefs. to helix/coil theory)

IT Amino acids, biological studies
RL: BPR (Biological process); BSU (Biological study, unclassified); BIOL (Biological study); PROC (Process)
(extending helix-coil **model** to include capping, side-chain interactions, 310 helixes, and N1, N2, and N3 preferences for amino acids)

IT Conformation
(protein; **detn**. of **.alpha.-helix** N1 energies after addn. of N1, N2, and N3 prefs. to helix/coil theory)

IT 56-41-7, L-Alanine, biological studies
RL: BPR (Biological process); BSU (Biological study, unclassified); BIOL (Biological study); PROC (Process)
(Ala has the highest preference for the N1 position of an **.alpha.-helix**)

L11 ANSWER 47 OF 67 CAPLUS COPYRIGHT 2002 ACS
AU Parodi, L.A.; Granatir, C.A.; Maggiora, G.M.
SO Comput. Appl. Biosci. (1994), 10(5), 527-35
CODEN: COABER; ISSN: 0266-7061
TI A consensus procedure for predicting the **location** of **.alpha.-helical** transmembrane segments in proteins
AB To aid in the development of three-dimensional models of membrane-bound

proteins, a consensus procedure for predicting **.alpha.-helical** transmembrane segments from amino acid sequences is presented. The algorithm combines the results of six individual prediction methods and some. . . developed which takes an input file contg. an amino acid sequence in one-letter code and outputs a list of the **.alpha.-helical** transmembrane segments predicted by the consensus algorithm.

ST protein **alpha helical** transmembrane segment localization; consensus procedure protein conformation

IT Conformation and Conformers
Simulation and **Modeling**, biological
(a consensus procedure for predicting the **location** of **.alpha.-helical** transmembrane segments in proteins)

IT Proteins, biological studies
RL: BOC (Biological occurrence); BIOL (Biological study); OCCU (Occurrence)
(**.alpha.-helical** transmembrane segment; a consensus procedure for predicting the **location** of **.alpha.-helical** transmembrane segments in proteins)

L11 ANSWER 52 OF 67 CAPLUS COPYRIGHT 2002 ACS

AU Zhou, Nian E.; Kay, Cyril M.; Sykes, Brian D.; Hodges, Robert S.

SO Biochemistry (1993), 32(24), 6190-7
CODEN: BICHAW; ISSN: 0006-2960

TI A single-stranded amphipathic **.alpha.-helix** in aqueous solution: Design, structural characterization, and its application for **determining .alpha.-helical** propensities of amino acids

AB To investigate the positional effect of **.alpha.-helical** propensities of amino acids in an amphipathic **.alpha.-helix**, an amphipathic **.alpha.-helical** model peptide (Ac-Glu-Ala-Glu-Lys-Ala-Ala-Lys-Glu-Ala-Glu-Lys-Ala-Ala-Lys-Glu-Ala-Glu-Lys-amide) was designed and characterized by CD and 2D-NMR spectroscopies. This peptide contains 65% **.alpha.-helical** structure in soln., and its monomeric mol. wt. in aq. soln. was **detd.** by size-exclusion chromatog. The independence of **.alpha.-helical** structure and stability on peptide concn. demonstrates that helix formation of this peptide is a monomol. process. To compare the. . . the absence and presence of TFE or urea. Apparently each amino acid has a different helix propensity when it is **located** in the hydrophobic face vs. hydrophilic face and the effect of substitution is more significant in the hydrophobic face. This single-stranded amphipathic **.alpha.-helical** peptide provides an appropriate model system to **det.** helix propensities of amino acids on both hydrophobic and hydrophilic faces.

ST peptide model **alpha helix** formation; amino acid helix formation peptide

IT Amino acids, properties
RL: PREP (Preparation)
(**.alpha.-helix** propensity of, **detn.** of, **model** peptide prepn. for)

IT Conformation and Conformers
(**.alpha.-helical**, amino acids propensity for, **detn.** of, peptide **model** for)

IT 140835-57-0P 149004-26-2P 149004-27-3P 149004-28-4P 149004-29-5P 149004-30-8P
RL: PREP (Preparation)
(prepn. and **.alpha.-helix** formation by, amino acid **.alpha.-helical** propensity **detn.** in relation to)

```
S (((protein or polypeptide or peptide) (2a) design) and angle)/clm,ab
         27273 PROTEIN/CLM
          7487 PROTEINS/CLM
         30747 PROTEIN/CLM
                 ((PROTEIN OR PROTEINS)/CLM)
         16382 PROTEIN/AB
          8414 PROTEINS/AB
         21566 PROTEIN/AB
                 ((PROTEIN OR PROTEINS)/AB)
         10274 POLYPEPTIDE/CLM
          1603 POLYPEPTIDES/CLM
         10959 POLYPEPTIDE/CLM
                 ((POLYPEPTIDE OR POLYPEPTIDES)/CLM)
          3744 POLYPEPTIDE/AB
          3696 POLYPEPTIDES/AB
          6400 POLYPEPTIDE/AB
                 ((POLYPEPTIDE OR POLYPEPTIDES)/AB)
         10143 PEPTIDE/CLM
          2622 PEPTIDES/CLM
         11410 PEPTIDE/CLM
                 ((PEPTIDE OR PEPTIDES)/CLM)
          4970 PEPTIDE/AB
          3900 PEPTIDES/AB
          7575 PEPTIDE/AB
                 ((PEPTIDE OR PEPTIDES)/AB)
        225274 DESIGN/CLM
          1402 DESIGNS/CLM
        225987 DESIGN/CLM
                 ((DESIGN OR DESIGNS)/CLM)
         24073 DESIGN/AB
          3518 DESIGNS/AB
         26596 DESIGN/AB
                 ((DESIGN OR DESIGNS)/AB)
            20 (PROTEIN OR POLYPEPTIDE OR PEPTIDE) (2A) DESIGN
        239246 ANGLE/CLM
         60956 ANGLES/CLM
        272554 ANGLE/CLM
                 ((ANGLE OR ANGLES)/CLM)
         75750 ANGLE/AB
         19421 ANGLES/AB
         90364 ANGLE/AB
                 ((ANGLE OR ANGLES)/AB)
L4           3 (((PROTEIN OR POLYPEPTIDE OR PEPTIDE) (2A) DESIGN) AND ANGLE)/CL
               M,AB

=> d bib,ab,clm

L4   ANSWER 1 OF 3  USPATFULL
AN     97:23331  USPATFULL
TI     Method of rational drug design based on ab initio computer simulation of
       conformational features of peptides
IN     Balaji, Vitukudi N., 1642 Orchard Wood Rd., Encinitas, CA, United States
       92024
       Singh, Chandra U., 1213 Orchard Glen Cir., Encinitas, CA, United States
       92024
PI     US 5612895              19970318
AI     US 1995-427118          19950421 (8)
RLI    Continuation of Ser. No. US 1994-223513, filed on 5 Apr 1994 which is a
       continuation of Ser. No. US 1990-628111, filed on 14 Dec 1990, now
       patented, Pat. No. US 5331573, issued on 19 Jul 1994
DT     Utility
FS     Granted
```

EXNAM Primary Examiner: Voeltz, Emanuel T.; Assistant Examiner: Choi, Kyle J.
LREP Seidman, StephanieBrown Martin Haller & McClain
CLMN Number of Claims: 19
ECL Exemplary Claim: 1
DRWN 46 Drawing Figure(s); 36 Drawing Page(s)
LN.CNT 3076

AB A method of rational drug **design** includes simulating **polypeptides** in a way that predicts the most probable secondary and/or tertiary structures of a polypeptide, e.g., an oligopeptide, without any presumptions as to the conformation of the underlying primary or secondary structure. The method involves computer simulation of the polypeptide, and more particularly simulating a real-size primary structure in an aqueous environment, shrinking the size of the polypeptide isobarically and isothermally, and expanding the simulated polypeptide to its real size in selected time periods. A useful set of tools, termed Balaji plots, energy conformational maps, and probability maps, assist in identifying those portions of the predicted peptide structure that are most flexible or most rigid. The rational design of novel compounds, useful as drugs, e.g., bioactive peptidomimetic compounds, and constrained analogs thereof, is thus made possible using the simulation methods and tools of the described invention.

CLM What is claimed is:

1. An ab initio computer-assisted method of predicting a stable tertiary structure of a peptide without any presumption regarding the underlying structural characteristics of the peptide, comprising: (a) simulating a real-size primary structure of a polypeptide in a solvent box, said primary structure comprising a plurality of amino acid residues linked together in a chain, each residue having .phi., .psi. **angles** associated therewith, said .phi., .psi. **angles** defining the relative **angle** of a first and second amide plane of said amino acid residue with a common C.sup..alpha. atom of said amino acid residue; (b) shrinking the size of the peptide isobarically and isothermally; (c) expanding the peptide to its real size in selected time periods; and (d) measuring the .phi., .psi. **angles** of the expanding amino acid residues.

2. The method of predicting a tertiary structure as set forth in claim 1 wherein step (d) further includes measuring the energy state of the expanding residues.

3. The method of predicting a tertiary structure as set forth in claim 1 wherein step (c) further includes expanding the peptide beyond its real size in selected time periods.

4. The method of predicting a tertiary structure as set forth in claim 3 further including analyzing the .phi., .psi. **angles** corresponding to at least two consecutive selected time periods in order to identify the differences therebetween, said differences being indicative of the rigidity of a particular amino acid residue within said polypeptide chain.

5. The method of predicting a tertiary structure as set forth in claim 4 wherein said step of analyzing the .phi., .psi. **angles** includes plotting the .phi., .psi. **angles** of the simulated peptide as a function of the residue.

6. The method of predicting a tertiary structure as set forth in claim 5 wherein the step of plotting the .phi., .psi. **angles** comprises (i) plotting the .phi. **angle** as the base of a wedge and the .psi. **angle** as the tip of a wedge along a first axis of a plot, said first axis having angular values marked thereon, with the tip of the wedge being aligned with the value of the .psi. **angle**

as indicated on the first axis, and the base of the wedge being aligned with the value of the .phi. **angle** as indicated on the first axis, and (ii) plotting a separate wedge for each amino acid residue along a second axis of said plot, said second axis being orthogonal to said first axis, said second axis having numerical values marked thereon, the wedge for a particular residue being aligned with the number of the residue as indicated on the second axis, whereby a wedge appears on said plot for each amino acid residue in said polypeptide chain, with the location of the base and tip of each wedge relative to said first axis indicating the .phi., .psi. **angles**, respectively, for the particular amino acid residue indicated by the location of the wedge relative to said second axis.

7. The method of predicting a tertiary structure as set forth in claim 1 wherein step (c) comprises expanding the amino acid residues of said polypeptide chain one at a time.

8. The method of predicting a tertiary structure as set forth in claim 7 further including biasing the expansion towards a structure predicted by known chemical and physical data.

9. The method of predicting a tertiary structure as set forth in claim 1 wherein step (c) comprises expanding the amino acid residues of said polypeptide chain simultaneously.

10. The method of predicting a tertiary structure as set forth in claim 9 further including biasing the expansion towards a structure predicted by known chemical and physical data.

11. A computer-assisted method for determining areas of flexibility and rigidness in a peptide, said peptide comprising a plurality of residues linked together in a chain, said method comprising the steps of: (a) electronically simulating said peptide in a fluid environment where the residue chain is free to move and fold as a result of natural molecular or electrical forces present in said residues; (b) measuring the .phi., .psi. **angles** associated with each residue of said simulated peptide at discrete time periods as said residue chain moves in said environment; (c) plotting the .phi., .psi. **angles** of the stimulated peptide as a function of the residue for a plurality of consecutive discrete time periods; and (d) determining the differences between the .phi., .psi. **angles** of corresponding residues of adjacent discrete time periods, whereby the relative flexibility or rigidness of a particular bond within said peptide is identified.

12. The method for determining areas of flexibility and rigidness in a peptide as set forth in claim 11 wherein step (a) includes: (i) shrinking the size of the simulated peptide isobarically and isothermally while in said fluid environment, and (ii) expanding the simulated peptide to its real size in discrete steps at each of said discrete time periods.

13. The method for determining areas of flexibility and rigidness in a peptide as set forth in claim 12 further including expanding the simulated peptide beyond its real size in discrete time periods.

14. The method for determining areas of flexibility and rigidness in a peptide as set forth in claim 13 wherein step (c) comprises (i) plotting the .phi. **angle** as the base of a wedge and the .psi. **angle** as the tip of a wedge along a first axis of a plot, said first axis having angular values marked thereon, with the tip of the wedge being aligned with the value of the .psi. **angle** as indicated on the first axis, and the base of the wedge being aligned

with the value of the .phi. **angle** as indicated on the first axis, and (ii) plotting a separate wedge for each amino acid residue along a second axis of said plot, said second axis being orthogonal to said first axis, said second axis having numerical values marked thereon, the wedge for a particular residue being aligned with the number of the residue as indicated on the second axis, whereby a wedge appears on said plot for each amino acid residue in said polypeptide chain, with the location of the base and tip of each wedge relative to said first axis indicating the .phi., .psi. **angles**, respectively, for the particular amino acid residue indicated by the location of the wedge relative to said second axis.

15. A system for determining areas of flexibility and rigidness in a peptide, said peptide comprising a plurality of residues linked together in a chain, said system comprising: (a) simulating means for simulating the structure of said peptide in a fluid environment where the residue chain is free to move and fold as a result of natural molecular or electrical forces present in said residues; (b) recording means for recording the .phi., .psi. **angles** associated with each residue of said simulated peptide at discrete time periods as said residue chain moves in said environment; (c) plotting means for plotting the .phi., .psi. **angles** of the simulated peptide as a function of the residue for a plurality of consecutive discrete time periods; and (d) analyzing means for analyzing the differences between the .phi., .psi. **angles** of corresponding residues of adjacent discrete time periods to identify the relative flexibility or rigidness of a particular bond within said peptide.

16. The system for determining areas of flexibility and rigidness in a peptide as set forth in claim 15 wherein said simulation means includes processing means for shrinking the size of the simulated peptide isobarically and isothermally while in said fluid environment, and expanding the simulated peptide to its real size in discrete steps at each of said discrete time periods.

17. The system for determining areas of flexibility and rigidness in a peptide as set forth in claim 16 wherein said processing means is for further expanding the simulated peptide beyond its real size in discrete time periods.

18. The system for determining areas of flexibility and rigidness in a peptide as set forth in claim 17 wherein said plotting means includes: means for plotting the .phi. **angle** as the base of a wedge and the .psi. **angle** as the tip of a wedge along a first axis of a plot, said first axis having angular values marked thereon, with the tip of the wedge being aligned with the value of the .psi. **angle** as indicated on the first axis, and the base of the wedge being aligned with the value of the .phi. **angle** as indicated on the first axis, and means for plotting a separate wedge for each amino acid residue along a second axis of said plot, said second axis being orthogonal to said first axis, said second axis having numerical values marked thereon, the wedge for a particular residue being aligned with the number of the residue as indicated on the second axis, whereby a wedge appears on said plot for each amino acid residue in said polypeptide chain, with the location of the base and tip of each wedge relative to said first axis indicating the .phi., .psi. **angles** , respectively, for the particular amino acid residue indicated by the location of the wedge relative to said second axis.

19. A method for generating biologically or pharmacologically active molecules, said method comprising: (i) determining the amino acid sequence of the hypervariable region of a monoclonal antibody having

biological or pharmacological activity, and (ii) producing a peptidomimetic compound based on the amino acid sequence of step (i), wherein said peptidomimetic compound substantially retains the biological or pharmacological activity of said monoclonal antibody.

=> d bib,ab,clm 2,3

L4 ANSWER 2 OF 3 USPATFULL
AN 96:109612 USPATFULL
TI Method of rational drug design based on AB initio computer simulation of conformational features of peptides
IN Balaji, Vitukudi N., 1642 Orchard Wood Rd., Encinitas, CA, United States 92024
Singh, Chandra U., 4 Thornhurst, San Antonio, TX, United States 78218
PI US 5579250 19961126
AI US 1995-427987 19950424 (8)
DCD 20110719
RLI Continuation of Ser. No. US 1994-223513, filed on 5 Apr 1994, now abandoned which is a continuation of Ser. No. US 1990-628111, filed on 14 Dec 1990, now patented, Pat. No. US 5331573, issued on 19 Jul 1994
DT Utility
FS Granted
EXNAM Primary Examiner: Trammell, James P.; Assistant Examiner: Choi, Kyle J.
LREP Seidman, Stephanie L.Brown Martin Haller & McClain
CLMN Number of Claims: 26
ECL Exemplary Claim: 1
DRWN 46 Drawing Figure(s); 36 Drawing Page(s)
LN.CNT 2539
AB A method of rational drug **design** includes simulating **polypeptides** in a way that predicts the most probable secondary and/or tertiary structures of a polypeptide, e.g., an oligopeptide, without any presumptions as to the conformation of the underlying primary or secondary structure. The method involves computer simulation of the polypeptide, and more particularly simulating a real-size primary structure in an aqueous environment, shrinking the size of the polypeptide isobarically and isothermally, and expanding the simulated polypeptide to its real size in selected time periods. A useful set of tools, termed Balaji plots, energy conformational maps, and probability maps, assist in identifying those portions of the predicted peptide structure that are most flexible or most rigid. The rational design of novel compounds, useful as drugs, e.g., bioactive peptidomimetic compounds, and constrained analogs thereof, is thus made possible using the simulation methods and tools of the described invention.
CLM What is claimed is:
1. A computer-assisted method of rational design of bioactive compounds, comprising: (a) electronically simulating and selecting the most probable conformations of a given polypeptide, wherein the most probable conformation of a given polypeptide is selected by an ab initio procedure, comprising: (i) simulating a real-size primary structure of a polypeptide in a solvent box, wherein the polypeptide comprises a plurality of amino acid residues linked together in a chain, each residue having .phi., .psi. **angles** associated therewith, said .phi., .psi. **angles** defining the relative **angle** of a first and second amide plane of said amino acid residue with a common C.sup..alpha. atom of said amino acid residue; (ii) shrinking the size of the polypeptide isobarically and isothermally; and (iii) expanding the polypeptide to its real size in selected time periods to determine an energetically most probable three-dimensional structure of the peptide; (b) designing a chemically modified analog that substantially mimics the energetically most probable three-dimensional structure of

the peptide; (c) chemically synthesizing the chemically modified analog of the peptide; and (d) evaluating the bioactivity of the synthesized chemically modified analog and selecting the analogs that exhibit bioactivity.

2. The method of rational design of claim 1, further comprising: (e) designing a peptidomimetic based on the conformation of the synthesized chemically modified analog.

3. The method of rational design of claim 1, wherein step (a) further comprises: (iv) recording the .phi., .psi. **angles** of the amino acid residues of the expanded peptide.

4. The method of rational design of claim 3 wherein step (a) further comprises: (v) determining the energy state of the expanded amino acid residues.

5. The method of rational design of claim 4, further comprising expanding the polypeptide beyond its real size in selected time increments, and recording the .phi., .psi. **angles** of the residues thus expanded.

6. The method of rational design of claim 4, wherein the step of simulating and selecting the most probable conformation of the peptides includes analyzing the recorded .phi., .psi. **angles** and energy states for each residue as a function of the time increments.

7. The method of rational design of claim 6, wherein the step of analyzing the recorded .phi., .psi. **angles** comprises plotting the .phi., .psi. **angles** for each amino acid residue of the simulated polypeptide as a function of the location of the residue within the polypeptide chain.

8. The method of rational design of claim 7 wherein the step of plotting the .phi., .psi. **angles** comprises: (i) plotting the .phi. **angle** as the base of a wedge and the .psi. **angle** as the tip of a wedge along a first axis of a plot, the first axis having angular values marked thereon, with the top of the wedge being aligned with the value of the .psi. **angle** as indicated on the first axis, and the base of the wedge being aligned with the value of the **angle** as indicated on the first axis, and (ii) plotting a separate wedge for each amino acid residue along a second axis of the plot, the second axis being orthogonal to the first axis, the second axis having numerical values marked thereon, the wedge for a particular residue being aligned with the number of the residue as indicated on the second axis, whereby a wedge appears on the plot for each amino acid residue in the polypeptide chain, with the location of he base and tip of each wedge relative to the first axis indicating the .phi., .psi. **angles**, respectively, for the particular amino acid residue indicated by the location of the wedge relative to the second axis.

9. The method of rational design of claim 6, further including analyzing the differences between the .phi., .psi. **angles** of corresponding residues at selected time increments to identify the relative flexibility or rigidness of a particular bond within the polypeptide.

10. The method of rational design of claim 9 wherein step (b) of designing and synthesizing a chemically modified analog of the selected peptide comprises identifying flexible portions of the polypeptide chain and replacing the flexible portions with bioisostere moieties.

11. The method of claim 10, wherein the flexible portions are identified by: (i) electronically simulating the peptide in a fluid environment where the residue chain is free to move and fold as a result of natural molecular or electrical forces present in the residues; (ii) measuring the .phi., .psi. **angles** associated with each residue of the simulated peptide at discrete time periods as the residue chain moves in the environment; (iii) plotting the .phi.,.psi. **angles** of the simulated peptide as a function of the residue for a plurality of consecutive discrete time periods; and (iv) determining the differences between the .phi., .psi. **angles** of corresponding residues of adjacent discrete time periods, whereby the relative flexibility or rigidness of a particular bond within the peptide is identified.

12. The method of claim 11, wherein step (i) comprises: (A) shrinking the size of the simulated peptide isobarically and isothermally while in the fluid environment; and (B) expanding the simulated peptide to its real size in discrete steps at each of the discrete time periods.

13. The method of claim 12, wherein the steps for determining areas of flexibility and rigidness in a peptide further comprise: (C) expanding the simulated peptide beyond its real size in discrete time periods.

14. The method of claim 13, wherein the step (iii) comprises: (A) plotting the .phi. **angle** as the base of a wedge and the .psi. **angle** as the tip of a wedge along a first axis of a plot, the first axis having angular values marked thereon, with the tip of the wedge being aligned with the value of the .psi. **angle** as indicated on the first axis, and the base of the wedge being aligned with the value of the .phi. **angle** as indicated on the first axis; and (B) plotting a separate wedge for each amino acid residue along a second axis of the plot, the second axis being orthogonal to the first axis, the second axis having numerical values marked thereon, the wedge for a particular residue being aligned with the number of the residue as indicated on the second axis, whereby a wedge appears on the plot for each amino acid residue in the polypeptide chain, with the location of the base and tip of each wedge relative to the first axis indicating the .phi., .psi. **angles**, respectively, for the particular amino acid residue indicated by the location for the wedge relative to the second axis.

15. The method of claim 14, wherein step (d) further comprises measuring the energy state of the expanding residues.

16. The method of claim 14, wherein step (c) further includes expanding the peptide beyond its real size in selected time periods.

17. The method of claim 16, further comprising analyzing the .phi., .psi. **angles** corresponding to at least two consecutive selected time periods in order to identify the differences therebetween, wherein the differences are indicative of the rigidity of a particular amino acid residue within the polypeptide chain.

18. The method of claim 17, wherein the step of analyzing the .phi., .psi. **angles** includes plotting the .phi., .psi. **angles** of the simulated peptide as a function of the residue.

19. The method of claim 18, wherein the step of plotting the .phi., .psi. **angles** comprises: (i) plotting the **angle** as the base of a wedge and the .phi. **angle** as the tip of a wedge along a first axis of a plot, the first axis having angular values marked thereon, with the tip of the wedge being aligned with the value of the .psi. **angle** as indicated on the first axis, and the

base of the wedge being aligned with the value of the .phi. **angle** as indicated on the first axis, and (ii) plotting a separate wedge for each amino acid residue along a second axis of the plot, the second axis being orthogonal to the first axis, the second axis having numerical values marked thereon, the wedge for a particular residue being aligned with the number of the residue as indicated on the second axis, whereby a wedge appears on the plot for each amino acid residue n the polypeptide chain, with the location of the base and tip of each wedge relative to the first axis indicating the .phi., .psi. **angles**, respectively, for the particular amino acid residue indicated by the location of the wedge relative to the second axis.

20. The method of claim 14, wherein step (c) comprises expanding the amino acid residues of the polypeptide chain one at a time.

21. The method of claim 20, further comprising biasing the expansion towards a structure predicted by known chemical and physical data.

22. The method of claim 14, wherein step (c) comprises expanding the amino acid residues of the polypeptide chain simultaneously.

23. The method of claim 22, further comprising biasing the expansion towards a structure predicted by known chemical and physical data.

24. A method for the design of a peptidomimetic or pharmacophore, the method comprising: (1) determining the energetically most probable tertiary structure of that portion of a pharmaceutically active compound that is responsible for the pharmacological action of the compound (2) producing a simulated, chemically modified peptide or peptidomimetic structure that substantially mimics the energetically most probable three-dimensional structure of the pharmaceutically active compound; (3) chemically synthesizing the chemically modified peptide or peptidomimetic structure; and (4) evaluating the bioactivity of the synthesized peptide or peptidomimetic structure.

25. The method of claim 24, wherein producing a simulated, chemically modified peptide or peptidomimetic structure that substantially mimics the energetically most probable three-dimensional structure of the pharmaceutically active compound is carried out by: (i) determining the .phi. and .psi. **angles** for each residue included in the energetically most probable conformation of the pharmaceutically active compound; (ii) comparing the .phi. and .psi. **angles** for each residue obtained in step (i) with the .phi. and .psi. **angles** for each residue of known polypeptide species, and (iii) substituting a chemically modified moiety for at least one of the residues of the pharmaceutically active compound, wherein the chemically modified moiety has .phi. and .psi. **angles** that are substantially similar to the .phi. and .psi. **angles** of the residue to be replaced.

26. The method of claim 24, wherein the energetically most probable tertiary structure of that portion of a pharmaceutically active compound which is responsible for the pharmacological action is determined by the method, comprising: (a) simulating a real-size primary structure of a polypeptide in a solvent box, the primary structure comprising a plurality of amino acid residues linked together in a chain, each residue having .phi., .psi. **angles** associated therewith, the .phi., .psi. **angles** defining the relative **angle** of a first and second amide plane of the amino acid residue with a common C.sup..alpha. atom of the amino acid residue; (b) shrinking the size of the peptide isobarically and isothermally; (c) expanding the peptide to its real size in selected time periods; and (d) measuring the .phi., .psi. **angles** of the expanding amino acid residues.

L4 ANSWER 3 OF 3 USPATFULL
AN 94:63030 USPATFULL
TI Method of design of compounds that mimic conformational features of selected peptides
IN Balaji, Vitukudi N., 1642 Orchard Wood Rd., Encinitas, CA, United States 92024
Singh, Chandra U., 1213 Orchard Glen Cir., Encinitas, CA, United States 92024
PI US 5331573 19940719
AI US 1990-628111 19901214 (7)
DT Utility
FS Granted
EXNAM Primary Examiner: Teska, Kevin J.
LREP Pretty, Schroeder, Brueggemann & Clark
CLMN Number of Claims: 3
ECL Exemplary Claim: 1
DRWN 46 Drawing Figure(s); 36 Drawing Page(s)
LN.CNT 2821
CAS INDEXING IS AVAILABLE FOR THIS PATENT.
AB A method of rational drug **design** includes simulating **polypeptides** in a way that predicts the most probable secondary and/or tertiary structures of a polypeptide, e.g., an oligopeptide, without any presumptions as to the conformation of the underlying primary or secondary structure. The method involves computer simulation of the polypeptide, and more particularly simulating a real-size primary structure in an aqueous environment, shrinking the size of the polypeptide isobarically and isothermally, and expanding the simulated polypeptide to its real size in selected time periods. A useful set of tools, termed Balaji plots, energy conformational maps, and probability maps, assist in identifying those portions of the predicted peptide structure that are most flexible or most rigid. The rational design of novel compounds, useful as drugs, e.g., bioactive peptidomimetic compounds, and constrained analogs thereof, is thus made possible using the simulation methods and tools of the described invention.
CLM What is claimed is:
1. A method for producing simulated, chemically modified peptide or peptidomimetic structure(s) which substantially mimic the energetically most probable three-dimensional structure of preselected less constrained polypeptide(s), said method comprising: (1) determining the .phi. and .psi. **angles** for each residue included in the preselected polypeptide; (2) comparing the .phi. and .psi. **angles** for each residue obtained in step (i) with the .phi. and .psi. **angles** for each residue of known polypeptide species; (3) substituting a chemically modified moiety for at least one of the residues of the preselected polypeptide to produce a chemically modified peptide or peptidomimetic structure, wherein said chemically modified moiety has .phi. and .psi. **angles** which are substantially similar to the .phi. and .psi. **angles** of the residue that is replaced; and (4) chemically synthesizing and testing the bioactivity of the chemically modified peptide or peptidomimetic structure.

2. The method of claim 1, wherein steps (1), (2), and (3) are repeated sequentially, beginning with a first residue of the preselected, less constrained polypeptide, so as to produce chemically modified analog(s) having a tertiary structure that substantially mimics the energetically most probable tertiary structure of the preselected, less constrained polypeptide(s).

3. A method for generating biologically or pharmacologically active molecules, comprising: (a) determining the amino acid sequence of the

hypervariable region of a monoclonal antibody having biological or phamacological activity, and (b) producing a peptidomimetic compound based on the amino acid sequence of step (a), wherein the peptidomimetic compound substantially retains the biological or pharmacological activity of said monoclonal antibody, wherein said peptidomimetic compound is produced by: (i) determining the energetically most probably .phi. and .psi. **angles** for each residue included in the hypervariable region of said monoclonal antibody, (ii) comparing the .phi. and .psi. **angles** for each residue obtained in step (i) with the .phi. and .psi. **angles** for each residue of known polypeptide species, and (iii) substituting a chemically modified moiety for at least one of the residues of the pharmaceutically active compound, wherein said chemically modified moiety has .phi. and .psi. **angles** which are substantially similar to the .phi. and .psi. **angles** of the residue to be repla